

Inference for a two-component mixture of symmetric distributions under log-concavity

Fadoua Balabdaoui and Charles R. Doss

Abstract

In this article, we revisit the problem of estimating the unknown zero-symmetric distribution in a two-component location mixture model, considered in previous works, now under the assumption that the zero-symmetric distribution has a log-concave density. When consistent estimators for the shift locations and mixing probability are used, we show that the nonparametric log-concave Maximum Likelihood estimator (MLE) of both the mixed density and that of the unknown zero-symmetric component are consistent in the Hellinger distance. In case the estimators for the shift locations and mixing probability are \sqrt{n} -consistent, we establish that these MLE's converge to the truth at the rate $n^{-2/5}$ in the L_1 distance. To estimate the shift locations and mixing probability, we use the estimators proposed by [Hunter et al. \(2007\)](#). The unknown zero-symmetric density is efficiently computed using the R package `logcondens.mode`.

1 Introduction

Let us assume that X_1, \dots, X_n are independent and identically distributed (i.i.d.) draws from a mixture distribution, with cumulative distribution function (cdf) G^0 given by

$$G^0(x) = \sum_{i=1}^k \pi_i^0 F_i^0(x), \quad x \in \mathbb{R}, \quad (1)$$

for some integer $k \geq 2$, where F_i^0 are cdfs, $\pi_i^0 \geq 0$, and $\sum_{i=1}^k \pi_i^0 = 1$. Such mixture distributions are very common in statistical modeling, in part because a variety of data generating frameworks lead to mixture models; for instance, one common approach to clustering problems leads to estimation of a mixture density ([Fraley and Raftery, 2002](#)). Another reason for this popularity is that they are very flexible and many distributions can be well approximated by some mixture model (see, e.g., [Everitt and Hand \(1981\)](#), [Titterton et al. \(1985\)](#), or [McLachlan and Peel \(2000\)](#)).

In this paper, we revisit the semi-parametric mixture model already studied by [Bordes et al. \(2006\)](#) and [Hunter et al. \(2007\)](#). In this model, it is assumed that the mixing distributions F_i , $1 \leq i \leq k$ in [\(1\)](#) are such that

$$F_i(x) = F^0(x - u_i^0)$$

for $u_i^0 \in \mathbb{R}$, $i = 1, \dots, k$, and F^0 is a distribution function restricted to be symmetric about 0, i.e. $F^0(-x) = 1 - F^0(x)$. This model was also studied more recently by [Butucea and Vandekerkhove \(2014\)](#). All of these authors have actually focused on the case $k = 2$:

$$G^0(x) = \pi^0 F^0(x - u_1^0) + (1 - \pi^0) F^0(x - u_2^0), \quad x \in \mathbb{R}. \quad (2)$$

This is still, in fact, a flexible model which is useful in many scenarios (see our data applications in [Section 6](#)). As the main goal is to estimate the mixing parameters and the mixing component F^0 , it is crucial to be assured that there exists a unique solution $(\pi^0, u_1^0, u_2^0, F^0)$ for a given G^0 determined by [\(2\)](#). [Bordes et al. \(2006\)](#) and [Hunter et al. \(2007\)](#) were able to establish that identifiability holds under some suitable conditions on the mixing parameters. Their result states that if $u_1^0 < u_2^0$ and $\pi^0 \notin \{0, 1/2, 1\}$, then G^0 given [\(2\)](#) is identifiable for any zero-symmetric distributions F^0 . Furthermore, the condition is necessary and sufficient since any distribution G^0 that is symmetric about its median clearly cannot be 2-identifiable; see [Theorem 2 of Hunter et al. \(2007\)](#).

After having shown identifiability, [Hunter et al. \(2007\)](#) put their focus on estimating $(\pi^0, u_1^0, u_2^0, F^0)$. They have shown that their estimator of the parametric component (π^0, u_1^0, u_2^0) is consistent and asymptotically normal. However, the obtained estimator of F^0 is not even guaranteed to have the properties of a genuine cdf (i.e., it is not necessarily nondecreasing). On the other hand, [Bordes et al. \(2006\)](#), [Butucea and Vandekerkhove \(2014\)](#), and [Chee and Wang \(2013\)](#) use a KDE approach to estimation of F^0 . The resulting estimators are proper distribution functions, but the procedures involve a model-selection procedure (cross-validation or Akaike or Bayesian information criterion) to choose the tuning parameter. The estimators of [Hunter et al. \(2007\)](#) and [Butucea and Vandekerkhove \(2014\)](#) for the mixture parameters are shown to converge weakly to a multivariate Gaussian at the parametric rate $n^{-1/2}$ under some regularity conditions on F^0 which are related to smoothness in the case of [Butucea and Vandekerkhove \(2014\)](#). [Bordes et al. \(2006\)](#) obtain also a convergence rate under smoothness assumptions, but their rate of convergence is much slower (of order $n^{-1/4+\alpha}$, for any $\alpha > 0$). [Bordes et al. \(2006\)](#) show that the same rates of convergence are inherited by their kernel estimator of F^0 in the supremum norm, under the assumption that the location parameters u_1^0 and u_2^0 are unknown. If F^0 is assumed to admit a density f^0 , then [Bordes et al. \(2006\)](#) provide only almost sure consistency in the supremum norm. For their kernel estimator, [Butucea and Vandekerkhove \(2014\)](#) obtain, for pointwise convergence, a rate of order $n^{-(2\beta-1)/(4\beta)}$ in the quadratic risk assuming smoothness of level $\beta > 1/2$ and assuming that the bandwidth is chosen optimally (the authors suggest using cross validation).

Hence, the proposed estimators of F^0 in the aforementioned works suffer various practical difficulties, including slow rates of convergence (or as-of-yet unknown rates) for it or its density, the estimator not being a proper cdf, or the need for model-selection procedures to choose a tuning parameter. Our goal in this paper is to circumvent those issues by constructing an estimator of the density f^0 which

- converges to the truth with a provably good convergence rate,
- can be efficiently computed,
- does not require a tuning parameter, and furthermore,
- is unimodal.

Unimodality is a natural constraint to enforce; when using a mixture model, it is somewhat unnatural to imagine a multimodal mixture component density. However, using unimodality involves some technical difficulties: enforcing unimodality on f^0 is not directly feasible, because the class of unimodal densities is too large and the MLE of a unimodal density fails to exist even in the simple one-dimensional setting (with no mixing). We propose instead to assume that f^0 satisfies the shape constraint of log-concavity (i.e., $\log f^0$ is concave).

Log-concave functions are always unimodal, and have been used to great success in nonparametric modeling. Unlike the class of unimodal densities, the log-concave class admits an MLE (Walter, 2002). Many papers have studied the log-concave MLE on \mathbb{R} or \mathbb{R}^d and much is already known about its large sample properties, both local and global; see e.g. Pal et al. (2007), Dümbgen and Rufibach (2009), Balabdaoui et al. (2009), Cule and Samworth (2010), Cule et al. (2010), Dümbgen et al. (2011), Schuhmacher and Dümbgen (2010), Chen and Samworth (2013), Doss and Wellner (2015), and Kim and Samworth (2014). Balabdaoui et al. (2013) studied asymptotics and confidence intervals of the discrete log-concave MLE of a probability mass function in the well- and misspecified settings. Dümbgen et al. (2010), Rufibach (2007) and Dümbgen and Rufibach (2011) study algorithms for computation of the MLE, allowing unequal weights to be assigned to the observations, an important feature of which we will take advantage.

In the present context, we need to consider the class of *zero-symmetric* log-concave densities on \mathbb{R} , which has not been considered before. To do so, we note that if f is zero-symmetric and log-concave on \mathbb{R} , then $f^+(t) := 2f(t)\mathbb{I}_{t \in [0, \infty)}$ is log-concave with mode at 0. Thus, through a simple transformation of the data, it can be shown that the original estimation problem is equivalent to maximizing the log-likelihood over the class of log-concave densities on $[0, \infty)$ with mode at 0. We can then compute the maximum of the log-likelihood easily by alternating between the EM algorithm (Dempster et al., 1977) and the active set algorithm provided in the R package **logcondens.mode** which computes the log-concave MLE with a fixed mode. We use the fact that the active set algorithm allows for unequal weights to be assigned to the data points: here, the weights assigned are proportional to the posterior probabilities from the EM algorithm.

We are able to show that the zero-symmetric log-concave MLE converges in probability to the true zero-symmetric log-concave component density in the Hellinger distance and in the supremum norm on sets of continuity of the true density. Furthermore, it can be shown that our estimator converges to the truth at the rate $n^{-2/5}$ in the L_1 -distance. Although the risk measure we use here is different from

the one considered by [Butucea and Vandekerckhove \(2014\)](#), it seems that the rate of convergence of our MLE, when the true mixture component is log-concave, is faster than that given in their Theorem 4 for their KDE when the smoothness parameter β satisfies $\beta \in (1/2, 5/2)$. Note for an estimator \hat{g}_n of g_0 in the direct density estimation problem based on i.i.d. observations from g_0 (as opposed to the mixture setting) when g_0 has smoothness β the optimal pointwise rate of convergence of $|\hat{g}_n(x_0) - g_0(x_0)|$ at a fixed point x_0 is $n^{-\beta/(2\beta+1)}$ ([Stone, 1980](#)). Note also that [Dümbgen and Rufibach \(2009\)](#) find a rate of convergence of $(\log n)^{\frac{\beta}{2\beta+1}} n^{-\frac{\beta}{2\beta+1}}$ in the uniform norm on compact sets for the log-concave MLE, in the direct density estimation problem, when the true density g_0 is log-concave and also lies in a Hölder class with smoothness β , i.e.

$$\begin{aligned} |g_0(x) - g_0(y)| &\leq L|x - y|, \text{ if } \beta = 1 \\ |g'_0(x) - g'_0(y)| &\leq L|x - y|^{\beta-1}, \text{ if } \beta > 1 \end{aligned}$$

for some $L > 0$. This rate is optimal for nonparametric estimation with smoothness β (the log factor being due to the supremum norm ([Khas'minskii, 1978](#))), and no bandwidth needs to be chosen.

We note that, although we refer to our estimator as the log-concave MLE, we do not use a “pure” maximum likelihood approach since we feed in other estimators of (π^0, u_1^0, u_2^0) to our likelihood, which we maximize to estimate f^0 and thus g^0 , the density of the mixed distribution G^0 . An alternative approach is to estimate both the parametric and nonparametric components simultaneously by maximum likelihood. However, there are many additional difficulties in that approach, due to the complicated non-concave nature of the log-likelihood function; see [Section 2](#).

We also note that we are not the first to use log-concavity in mixture modeling; [Chang and Walther \(2007\)](#) and [Eilers and Borgdorff \(2007\)](#) consider univariate mixtures of log-concave densities, and [Cule et al. \(2010\)](#) consider multivariate mixtures of log-concave densities. However, in none of those settings was symmetry imposed, perhaps because the authors were not worried about the (often fundamental) question of identifiability. Thus, their work does not directly apply in our setting.

The paper will be structured as follows. In [Section 2](#) we establish existence of the MLE and provide a necessary condition for a candidate to be equal to the estimator. In [Section 3](#), we establish consistency in the Hellinger distance. This implies other forms of consistency by the results of [Cule and Samworth \(2010\)](#). The techniques we used are re-adapted from [Pal et al. \(2007\)](#), [Cule and Samworth \(2010\)](#) and [Schuhmacher and Dümbgen \(2010\)](#) to deal with the additional difficulties of a mixture model. In [Section 4](#), we find that the MLEs of f^0 and g^0 converge to the truth at a rate of order $n^{-2/5}$. In [Section 5](#), we develop a likelihood ratio procedure based on our estimator in the problem of testing absence of mixing. We also consider the problem of clustering where we use the estimators of the posterior probabilities obtained via our log-concave MLE. In both problems, we compare our method to alternative or existing approaches. In [Section 6](#), we present two data applications. [Section 7](#) gathers some conclusions. Proofs and technical details can be found in

the online supplementary material.

2 The model and estimation via Maximum likelihood

Let X_1, \dots, X_n to be n independent observations assumed to come from the location mixture with cdf G^0 which we now assume has a density, given by

$$g^0(x) = \pi^0 f^0(x - u_1^0) + (1 - \pi^0) f^0(x - u_2^0), \quad (3)$$

for some $\pi^0 \in (0, 1) \setminus \{1/2\}$, $u_1^0, u_2^0 \in \mathbb{R}$ such that $u_1^0 \neq u_2^0$. We assume that f^0 is a zero-symmetric log-concave density, i.e. $f^0 \in \mathcal{S}\mathcal{L}\mathcal{C}_1$ where

$$\mathcal{S}\mathcal{L}\mathcal{C}_1 := \mathcal{S}\mathcal{L}\mathcal{C} \cap \left\{ f : \int_{\mathbb{R}} f(u) du = 1 \right\}, \text{ and } \mathcal{S}\mathcal{L}\mathcal{C} := \left\{ e^\psi : \psi \in \mathcal{S}\mathcal{C} \right\},$$

and $\mathcal{S}\mathcal{C}$ is the class of concave functions on \mathbb{R} that are upper semi-continuous (“closed”) and proper (Rockafellar, 1970), and satisfy $\psi(x) = \psi(-x)$. The upper semi-continuity condition is made only for the purpose of uniqueness. Then

$$L(\pi, u_1, u_2, f) := \sum_{j=1}^n \log \left(\pi f(X_j - u_1) + (1 - \pi) f(X_j - u_2) \right). \quad (4)$$

is the log-likelihood in this problem. In the case of estimation of a log-concave density on \mathbb{R} , the log-likelihood is a concave function (Pal et al. (2007), Rufibach (2007), Dümbgen and Rufibach (2009)). However, Dümbgen et al. (2010) study a semiparametric model incorporating log-concavity and find a non-concave likelihood; see their Section 3.3 including a plot on page 18. Unfortunately, our objective function L is also far from concave. Consider order statistics $X_{(1)}, \dots, X_{(n)}$, a fixed $\pi \in (0, 1) \setminus \{1/2\}$, and (zero-symmetric log-concave) f with support given by $\text{supp}(f) = [-s, s]$ and $\inf_{x \in \text{supp}(f)} f(x) > 0$. Assume $u_1 < u_2$ are such that $[u_1 - s, u_1 + s] \cup [u_2 - s, u_2 + s] \supset [X_{(1)}, X_{(n)}]$ so that $L(\pi, u_1, u_2, f) > -\infty$. Let $j := \min \{i : X_{(i)} > u_2 - s\}$ be the index of the smallest order statistic contained in the support of the second component, and let $\tilde{\delta} := X_{(j)} - (u_2 - s)$. Then not only does $L(\pi, u_1, \cdot, f)$ fail to be concave, but it is in fact discontinuous at $u_2 + \tilde{\delta}$.

We now describe our estimation approach. Let $\tilde{\pi}_n, \tilde{u}_{1,n}, \tilde{u}_{2,n}$ be estimators of π, u_1, u_2 , where we assume $\tilde{\pi}_n \in (0, 1) \setminus \{1/2\}$ and $\tilde{u}_{1,n} < \tilde{u}_{2,n}$. We will generally think of these estimators as being \sqrt{n} -consistent. We will then consider maximizing the log-likelihood

$$f \mapsto \sum_{j=1}^n \log \left(\tilde{\pi}_n f(X_j - \tilde{u}_{1,n}) + (1 - \tilde{\pi}_n) f(X_j - \tilde{u}_{2,n}) \right) \quad (5)$$

over $\mathcal{S}\mathcal{L}\mathcal{C}_1$. Using the Lagrange penalty term introduced by Silverman (1982), this is equivalent to maximizing the criterion Φ_n defined as

$$\Phi_n(\psi) = \frac{1}{n} \sum_{j=1}^n \log \left[\tilde{\pi}_n e^{\psi(X_j - \tilde{u}_{1,n})} + (1 - \tilde{\pi}_n) e^{\psi(X_j - \tilde{u}_{2,n})} \right] - \int_{\mathbb{R}} e^{\psi(x)} dx$$

over \mathcal{SC} . We will abusively use the term MLE for our estimators of f^0 and g^0 despite the fact that the mixing parameters (π, u_1, u_2) are not a part of the space over which the likelihood is maximized. In the next proposition we establish existence of the MLE, and describe its nature.

Proposition 2.1 *The criterion Φ_n admits a maximizer $\hat{\psi}_n$. Furthermore, the following holds true almost surely, letting $\hat{f}_n = e^{\hat{\psi}_n}$.*

- \hat{f}_n is in \mathcal{SLC}_1 .
- For $i = 1, \dots, n$, let

$$Z_{2i-1} = |X_i - \tilde{u}_{1,n}| \text{ and } Z_{2i} = |X_i - \tilde{u}_{2,n}|. \quad (6)$$

Then, on $[0, \infty)$ the MLE $\hat{\psi}_n$ changes slope only at points belonging to the set

$$\{Z_1, Z_2, \dots, Z_{2n-1}, Z_{2n}\}.$$

Furthermore, $\hat{\psi}'_n(0) = 0$, and $\hat{\psi}_n(x) = -\infty$ if and only if $x \notin [-Z_{(2n)}, Z_{(2n)}]$ where $Z_{(2n)}$ is the largest order statistic of $Z_1, Z_2, \dots, Z_{2n-1}, Z_{2n}$.

The MLE of f will be denoted by \hat{f}_n throughout, and that of g by \hat{g}_n . In the following, we give a necessary condition for a log-concave function $f = \exp(\psi)$ to be the MLE.

Proposition 2.2 is interesting to compare with the characterization of [Dümbgen and Rufibach \(2009\)](#) for the log-concave MLE. The result is also useful in combination with the EM-algorithm described below as its non-fulfillment indicates that convergence is not yet reached.

Proposition 2.2 *Let ψ be a zero-symmetric concave function on \mathbb{R} such that $\psi(x) = -\infty$ if and only if $x \notin [-Z_{(2n)}, Z_{(2n)}]$ where $Z_{(2n)}$ is defined in Proposition 2.1, and $\psi'(0) = 0$. If $\exp(\psi) = \hat{f}_n$ is the MLE, then for any real zero-symmetric function Δ such that $\psi + \epsilon\Delta \in \mathcal{SC}$ for some $\epsilon > 0$ we have that*

$$\frac{1}{n} \sum_{i=1}^n \left\{ \hat{p}_n(X_i) \Delta(X_i - \tilde{u}_{1,n}) + (1 - \hat{p}_n(X_i)) \Delta(X_i - \tilde{u}_{2,n}) \right\} \leq \int_{\mathbb{R}} \hat{f}_n(x) \Delta(x) dx \quad (7)$$

where

$$\hat{p}_n(X_i) = \frac{\tilde{\pi}_n \hat{f}_n(X_i - \tilde{u}_{1,n})}{\tilde{\pi}_n \hat{f}_n(X_i - \tilde{u}_{1,n}) + (1 - \tilde{\pi}_n) \hat{f}_n(X_i - \tilde{u}_{2,n})} = \frac{\tilde{\pi}_n \hat{f}_n(X_i - \tilde{u}_{1,n})}{\hat{g}_n(X_i)}, \quad (8)$$

for $i = 1, \dots, n$.

Next, we give the condition in (7) under an alternative form. [Dümbgen and Rufibach \(2009\)](#) shows that the log-concave MLE is uniquely characterized by the fact that the first integral of the cdf of the MLE stays below the first integral of the empirical distribution, while touching it exactly at the points where the logarithm of the MLE changes slope. To derive a related result, let $\hat{\mathbb{F}}_n$ denote the cdf of the discrete distribution putting mass $\hat{p}_n(X_i)/n$ at Z_{2i-1} and $(1 - \hat{p}_n(X_i))/n$ at Z_{2i} for $i = 1, \dots, n$, where $\hat{p}_n(X_i)$ was defined in (8) and Z_i was defined in (6). That is,

$$\hat{\mathbb{F}}_n = \frac{1}{n} \sum_{i=1}^n (\hat{p}_n(X_i) \delta_{Z_{2i-1}} + (1 - \hat{p}_n(X_i)) \delta_{Z_{2i}})$$

where $\delta_x(t) = 1_{\{[x, \infty)\}}(t)$. Let $\hat{f}_n^+(x) = 2\hat{f}_n(x)1_{x \in [0, \infty)}$, $\hat{\psi}_n^+ = \log(\hat{f}_n^+)$ and let \hat{F}_n^+ be the cdf of \hat{f}_n^+ .

Proposition 2.3 *If \hat{f}_n is the MLE of the component $f^0 \in \mathcal{S}\mathcal{L}\mathcal{C}_1$ then*

$$\int_0^z \hat{F}_n^+(x) dx \begin{cases} \leq \int_0^z \hat{\mathbb{F}}_n(x) dx, & \text{for } z \in [0, Z_{(2n)}] \\ = \int_0^z \hat{\mathbb{F}}_n(x) dx, & \text{if } \hat{\psi}_n^+(z-) > \hat{\psi}_n^+(z+). \end{cases} \quad (9)$$

3 Consistency

The main result of this section is to establish consistency in the Hellinger distance of the MLEs \hat{g}_n and \hat{f}_n as $n \rightarrow \infty$, where the Hellinger distance is defined by

$$H(p, q) := \sqrt{(1/2) \int \left(\sqrt{p(x)} - \sqrt{q(x)} \right)^2 dx}.$$

We will also find consistency for \hat{f}_n in certain exponentially weighted metrics. Our approach to the problem follows the idea of [Pal et al. \(2007\)](#) and [Dümbgen and Rufibach \(2009\)](#) but will require handling carefully the extra complexity induced by the mixture. As in [Pal et al. \(2007\)](#), [Cule and Samworth \(2010\)](#) and [Schuhmacher and Dümbgen \(2010\)](#), we will first need to show that the MLE of the mixed density and hence the MLE of the log-concave component are bounded. Here, the claimed boundedness will be only in probability, which is weaker than the almost sure boundedness proved in the aforementioned articles. Those articles, however, were able to take advantage of the fact that the level sets of a bounded unimodal function are convex and compact; such a statement does not hold if we consider a mixture of two unimodal functions instead of a single unimodal function, even if the two components are log-concave. So, instead of studying how the empirical distribution behaves over the class of compact intervals, we will instead need to study its behavior over more complicated classes of functions. This is what is done in Propositions [A.2](#) and [A.3](#).

Theorem 3.1 *Let g^0 be as in (3) and \hat{g}_n be the MLE of g^0 . Then we have that*

$$H(\hat{g}_n, g^0) = o_p(1).$$

Consistency of the log-concave component, \hat{f}_n , follows now from Theorem 3.1.

Corollary 3.1 *Let f^0 denote again the true log-concave zero-symmetric density. Then,*

$$H(\hat{f}_n, f^0) = o_p(1),$$

and for any $a \in (0, a_0)$ such that $f^0(x) \leq \exp(-a_0x + b)$ for some $b \in \mathbb{R}$, then

$$\int_{\mathbb{R}} e^{at} |\hat{f}_n(t) - f^0(t)| = o_p(1),$$

and

$$\sup_{t \in [-A, A]} e^{at} |\hat{f}_n(t) - f^0(t)| = o_p(1).$$

on any continuity set $[-A, A]$ of f^0 , where A may be ∞ if f^0 is continuous on all of \mathbb{R} .

4 Rates of convergence

In this section, we aim at refining the convergence result obtained in the previous section to attain a rate of convergence for both \hat{f}_n and \hat{g}_n in the L_1 distance. To this goal, we need first to recall some definitions from empirical processes theory. Given a class of functions \mathcal{F} , the bracketing number of \mathcal{F} under some distance $\|\cdot\|$ is defined as

$$N_{[\cdot]}(\epsilon, \mathcal{F}, \|\cdot\|) = \min \left\{ k : \exists \underline{f}_1, \bar{f}_1, \dots, \underline{f}_k, \bar{f}_k \text{ s.t. } \| \underline{f}_i - \bar{f}_j \| \leq \epsilon, \mathcal{F} \subset \cup_{i=1}^k [\underline{f}_i, \bar{f}_i] \right\}$$

where $[l, u] = \{f : l \leq f \leq u\}$. In this section, we refine the consistency result above by deriving the rate of convergence of the MLE's \hat{g}_n and \hat{f}_n of the mixed density and the zero-symmetric log-concave component respectively.

For fixed $M > 0$, $a_0 < b_0$ and $\delta \in (0, (b_0 - a_0)/2)$, consider the class of functions

$$\mathcal{G} = \left\{ \lambda f(\cdot - a) + (1 - \lambda)f(\cdot - b), f \in \mathcal{SLC}, f(0) \in [1/M, M], \lambda \in [0, 1], \right. \\ \left. (a, b) \in [a_0 - \delta, a_0 + \delta] \times [b_0 - \delta, b_0 + \delta] \right\}.$$

Here, the parameters a_0 and b_0 play the role of the true location shifts u_1^0 and u_2^0 . Consistency of the estimates $\check{u}_{1,n}$ and $\check{u}_{2,n}$ ensures that they stay within distance 2δ from the truth with increasing probability. Also, uniform consistency of the log-concave MLE, \hat{f}_n , on continuity sets of f^0 implies consistency at the point 0 (the common mode of f^0 and \hat{f}_n). Thus, we can find $M > 0$ such that $\hat{f}_n(0) \in [1/M, M]$ with increasing probability. The following proposition gives a bound on the bracketing entropy for the class \mathcal{G} .

Proposition 4.1 *For $\epsilon \in (0, \epsilon_0]$, we have that*

$$\log N_{[\]}(\epsilon, \mathcal{G}, H) \lesssim \frac{1}{\sqrt{\epsilon}}$$

where ϵ_0 and \lesssim depend only on a_0, b_0, δ and M .

Now, we are ready to state our main theorem. We find a rate of convergence of at least $n^{-2/5}$ in the L_1 norm, both for \hat{f}_n and \hat{g}_n . Although we consider L_1 and Butucea and Vandekerkhove (2014) consider L_2 distance, the rate of Theorem 4.1 is an improvement over the corresponding L_2 rate $n^{-(2\beta-1)/(4\beta)}$ of Butucea and Vandekerkhove (2014) whenever $\beta < 5/2$. (Note that (log-)concave functions are Lebesgue-almost-everywhere twice differentiable by Alexandrov's theorem (Niculescu and Persson, 2006), so roughly correspond to β being 2 or larger.)

Theorem 4.1 *Let \hat{f}_n and \hat{g}_n be again the MLE's of the zero-symmetric log-concave component and mixed density respectively. If $\sqrt{n}(\check{u}_{1,n} - u_1^0) = O_p(1)$, $\sqrt{n}(\check{u}_{2,n} - u_1^0) = O_p(1)$, and $\sqrt{n}(\check{\pi}_n - \pi^0) = O_p(1)$, then*

$$L_1(\hat{f}_n, f^0) = O_p(n^{-2/5}), \text{ and } L_1(\hat{g}_n, g^0) = O_p(n^{-2/5})$$

where $L_1(d_1, d_2) = \int_{\mathbb{R}} |d_1(x) - d_2(x)| dx$.

To illustrate the theory, a simulated example is given in Figure 1. The true zero-symmetric component f^0 is taken to be the density of a standard Gaussian with mixing probability $\pi^0 = 1/3$ and shift locations $u_1^0 = 0$ and $u_2^0 = 4$. The plot on the left (right) shows our MLE of g^0 (f^0) based on a sample of size $n = 500$. The bullets in the right plot depict the knot points of the zero-symmetric log-concave MLE, that is, the points where the logarithm of the log-concave MLE changes its slope.

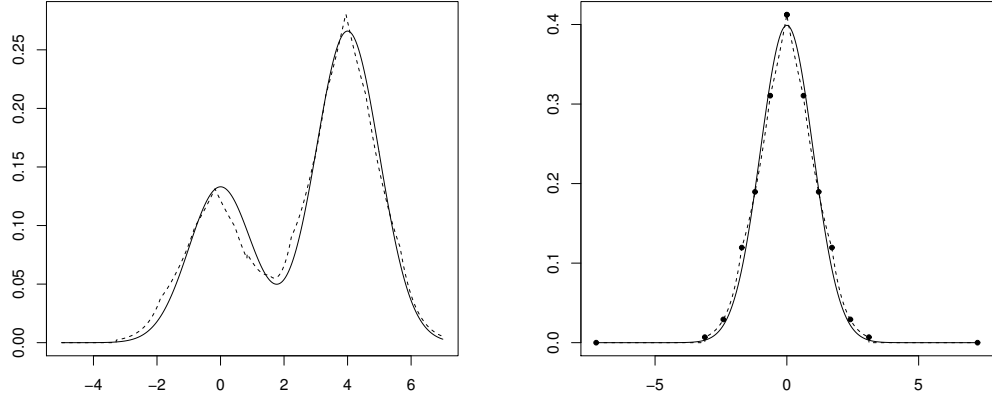


Figure 1: Left: plot of the true density of $(1/3)\mathcal{N}(0, 1) + (2/3)\mathcal{N}(4, 1)$ (solid line) and its log-concave estimator \hat{g}_n (dotted line). Right: plot of the density of $\mathcal{N}(0, 1)$ and its zero-symmetric log-concave MLE \hat{f}_n . The MLE was based on $n = 500$ independent data drawn from the mixture density $(1/3)\mathcal{N}(0, 1) + (2/3)\mathcal{N}(4, 1)$.

5 Testing and clustering

Bordes et al. (2006), Hunter et al. (2007) and Butucea and Vandekerkhove (2014) propose three different ways of estimating the mixture parameters π^0, u_1^0 and u_2^0 . As we are interested here in \sqrt{n} -consistent estimators of these parameters, we prefer the work by Hunter et al. (2007) and Butucea and Vandekerkhove (2014). Also, due to some numerical instabilities encountered when computing the estimators proposed by Butucea and Vandekerkhove (2014), we adopt the approach of Hunter et al. (2007) which has been already implemented in R; one could either use the code posted at <http://www.stat.psu.edu/~dhunter/code> or the function in the **mixtools** package. The latter option was kindly brought to the attention of the first author by David Hunter in a private communication.

Once the estimates $\tilde{\pi}_n, \tilde{u}_{1,n}$ and $\tilde{u}_{2,n}$ of π^0, u_1^0 and u_2^0 are computed, we maximize

$$f^+ \mapsto \frac{1}{n} \sum_{i=1}^n \log (\tilde{\pi}_n f^+(Z_{2i-1}) + (1 - \tilde{\pi}_n) f^+(Z_{2i}))$$

where $f^+ = 2f1_{[0,\infty)}$, $Z_{2i-1} = |X_i - \tilde{u}_{1,n}|$ and $Z_{2i} = |X_i - \tilde{u}_{2,n}|$ for $i = 1, \dots, n$. This is equivalent to maximizing (5), which we have shown to admit a maximizer. Since the log-likelihood is not concave and it is not clear how to maximize it directly, we will appeal to the EM algorithm (Dempster et al., 1977). Although we are fixing the parameters π, u_1 , and u_2 we may still introduce the standard-in-mixture-models complete data of (X_i, Δ_i) , where $\Delta_i \sim \text{Bernoulli}(\pi)$ and $X_i | \{\Delta_i = 1\} \sim f^0(\cdot - \tilde{u}_{1,n})$

and $X_i | \{\Delta_i = 0\} \sim f^0(\cdot - \check{u}_{2,n})$. An iteration of the EM algorithm in this setup is then, given an estimate $\hat{f}_n^{+, (r)}$, to compute

$$\hat{f}_n^{+, (r+1)} := \operatorname{argmax}_{f^+} \sum_{i=1}^n \left\{ \hat{p}_{n,i}^{(r)} \log f^+(Z_{2i-1}) + (1 - \hat{p}_{n,i}^{(r)}) \log f^+(Z_{2i}) \right\} \quad (10)$$

where the argmax is over log-concave densities on $[0, \infty)$ with mode at 0, and

$$\hat{p}_{n,i}^{(r)} = \frac{\check{\pi}_n \hat{f}_n^{+, (r)}(Z_{2i-1})}{\check{\pi}_n \hat{f}_n^{+, (r)}(Z_{2i-1}) + (1 - \check{\pi}_n) \hat{f}_n^{+, (r)}(Z_{2i})}$$

is the conditional expectation of Δ_i given X_i . To initialize the EM algorithm, we start with the density of a centered Gaussian distribution with variance equal to the estimate given in formula (11) of [Hunter et al. \(2007\)](#) for the true variance of the zero-symmetric component, that is $\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 - \check{\pi}_n(1 - \check{\pi}_n)(\check{u}_{2,n} - \check{u}_{1,n})^2$, or 1 if this estimate is negative (this may occur for moderate sample sizes). The argmax in (10) can be computed by the R package **logcondens.mode**.

5.1 Testing the absence of mixing

Recall that the mixing model we consider in this paper is given by

$$g^0 = \pi^0 f^0(\cdot - u_1^0) + (1 - \pi^0) f^0(\cdot - u_2^0)$$

with f^0 a log-concave zero-symmetric density on \mathbb{R} , $\pi^0 \notin \{0, 1/2, 1\}$ and $u_1^0 < u_2^0$. We now use our log-concave MLE to test for the absence of mixing, i.e. to test for the null hypothesis that $u_1^0 = u_2^0$, against the alternative that $u_1^0 \neq u_2^0$ and $\pi^0 \neq 1/2$ under the assumption that f^0 is zero-symmetric and log-concave.

To test for mixing we consider the likelihood ratio statistic. Under the null hypothesis, we take the estimator of the true density to be equal to the log-concave MLE which is symmetric around the median of the data. If \hat{g}_n^0 denotes this estimator, then our test statistic is given by

$$\Lambda_n = \frac{\prod_{i=1}^n \hat{g}_n(X_i)}{\prod_{i=1}^n \hat{g}_n^0(X_i)}. \quad (11)$$

The null hypothesis is then rejected when Λ_n is too large. We use the null hypothesis estimator to find critical values; that is, we bootstrap from the symmetric log-concave estimator \hat{g}_n^0 . The critical values of Λ_n are then computed in the usual way: based on the bootstrapped samples from \hat{g}_n^0 , we compute the estimators of the mixing probability and mixture locations and the corresponding MLE \hat{g}_n . The order statistics of the bootstrapped values of the likelihood ratio are then obtained to compute upper empirical quantiles of a given order. We also compare our test for mixing (hereafter referred to as the LR test) to the following procedures:

- the naive symmetric bootstrap (NSBS): we re-sample with replacement n random variables Z_1^*, \dots, Z_n^* from $\{\pm|X_1 - \hat{m}_n|, \dots, \pm|X_n - \hat{m}_n|\}$ with \hat{m}_n the median of X_1, \dots, X_n and set $X_i^* = \hat{m}_n + Z_i^*, i = 1, \dots, n$. Then, the bootstrapped estimators of the location mixture of Hunter et al. (2007), u_1^* and u_2^* , are computed based on $Y_i^*, i = 1, \dots, n$. We repeat this procedure B times and compute the empirical $(1 - \alpha)$ -quantile of the distribution of $u_2^* - u_1^*$. The null hypothesis is rejected if the observed $\check{u}_{2,n} - \check{u}_{1,n}$ is larger than this quantile.
- the naive symmetric bootstrap based on symmetric kernel density estimation (NSBSKDE): the method is similar to the one described above except that a standard kernel density estimator is fitted to $\hat{m}_n \pm |X_i - \hat{m}_n|$ and X_1^*, \dots, X_n^* are now drawn from the fitted estimator at each bootstrap iteration.
- the likelihood ratio based on symmetric kernel density estimation (LRSKDE): two kernel density estimators are computed, one under the full model, that is, based on X_1, \dots, X_n , and one under the null model, that is, based on $\hat{m}_n \pm |X_i - \hat{m}_n|$. The likelihood ratio of these estimators is then computed. Bootstrap samples are obtained by simulation from the kernel estimator under the null hypothesis and then the empirical $(1 - \alpha)$ -quantile of the likelihood ratio is thereby computed. The null hypothesis is rejected if the observed likelihood ratio is larger than this quantile.

Note that the NSBS provides a comparison procedure not based on density estimation of the components. In assessing the power, we take the true zero-symmetric component f^0 to be one of the following distributions: (1) a standard Gaussian, (2) a double exponential, and (3) a uniform on $[-1, 1]$. Also, we take the true parameters to be $\pi^0 \in \{0.20, 0.40\}$ and $(u_1^0, u_2^0) \in \{(0, 0), (0, 1), (0, 3)\}$. We give the estimated probability of rejecting the null hypothesis based on $R = 500$ replications with $B = 49$ bootstrap samples in Table 1, for $n = 250$. The simulation results show the LR and LRSKDE tests are both outperforming the NSBS and NSBSKDE with power nearly equal or equal to 1 for the well-separated mixtures. However, all the considered tests seem to have a level larger than the specified level $\alpha = 0.1$ for the uniform distribution. Further simulations, which we do not report here, show that this improves when the sample size is increased to $n = 500$. Note that the mixtures with mixture probability $\pi^0 = .4$ are more difficult to distinguish than those with $\pi^0 = .2$. This is to be expected as the former mixtures are close to being symmetric around the mid-point $(u_1^0 + u_2^0)/2$.

It would be interesting to know whether the level of our testing procedure based on the bootstrapped likelihood ratio test equals the theoretical level. The problem is however far from being trivial. Deriving the asymptotic level for example would require establishing the limit distribution of our statistic under the null hypothesis and also showing that it admits a continuous cumulative distribution function. Establishing such results requires a thorough study of the global asymptotics of the

log-concave MLE. As this is outside the scope of this paper, the question remains open.

5.2 Gaussian versus symmetric log-concave clustering

We now consider the problem of clustering, i.e., of assigning to each observation in a dataset a label without being given any “training” labels. We will assume that the data can be clustered into two groups, which we will do by fitting the two-component mixture (2) and assigning a label to an observation X based on whether our estimate of the posterior probability

$$\frac{\pi^0 f^0(X - u_1^0)}{\pi^0 f^0(X - u_1^0) + (1 - \pi^0) f^0(X - u_2^0)} \quad (12)$$

is greater than $1/2$ or not.

We fit the mixture three different ways. In the first basic approach, labeled “G”, we maximize the likelihood (5) under the assumption that the component f is a normal density. We use the EM algorithm to maximize the likelihood. Our next two approaches both use the method of Hunter et al. (2007) to estimate the mixture components u_1^0, u_2^0 , and π^0 . Then we either fit the components using a Gaussian density (denoted “HG”), with variance estimate also given by Hunter et al. (2007), or we use the symmetric log-concave density estimator (denoted “SLC”) for the components. The fourth approach is based on the estimators of Hunter et al. (2007) and the kernel density estimator based on the inversion formula given in (9) by Bordes et al. (2006) where we truncate the infinite sum at some large integer $K > 0$. Precisely, let \bar{g}_n be a standard density estimator of the mixed density g^0 . Then, the KDE of f^0 we use is given by $\max(1/2(\bar{f}_n(x) + \bar{f}(-x)), 0)$ where

$$\bar{f}_n(x) = \sum_{k=0}^K \left(\frac{-\tilde{\pi}_n}{1 - \tilde{\pi}_n} \right)^k \bar{g}_n(x + \tilde{u}_{2,n} + k(\tilde{u}_{2,n} - \tilde{u}_{1,n})).$$

We should note that this formula is only valid, when $\tilde{\pi}_n < 1/2$. Hence, $1 - \tilde{\pi}_n$ and $(\tilde{u}_{2,n}, \tilde{u}_{1,n})$ should replace $\tilde{\pi}_n$ and $(\tilde{u}_{1,n}, \tilde{u}_{2,n})$ When $\tilde{\pi}_n \geq 1/2$.

We record the average missclassification count when the true density is one of the densities in the left column of Table 2. In all cases the number of replications is $R = 5000$, the sample size is $n = 500$, $u_2^0 - u_1^0 = 1$ and $\pi^0 = .2$.

The performances of the four approaches were then compared, and the results are reported in Table 2. The KDE approach does clearly worse than the three other methods. The SLC outperforms HG by 16% when the true density is $\mathcal{U}[-1, 1]$. In the other cases, they perform similarly. All four methods define the two cluster regions by dividing the real line into two half-lines. The HG and SLC methods have the same mixture components so the shape of the component density estimates have to be dramatically different (e.g., uniform instead of normal) in order to noticeably change the results; note this somewhat deceiving outcome is not totally in contradiction with the finding of Cule et al. (2010) about the performance of their

Distribution	π^0	Test	$u_2^0 - u_1^0 =$		
			0	1	3
$\mathcal{N}(0, 1)$.2	LR	0.11	0.15	1.00
		NSBS	0.06	0.07	0.34
		NSBSKDE	0.07	0.08	0.40
		LRSKDE	0.11	0.11	1.00
	.4	LR	*	0.14	0.86
		NSBS	*	0.05	0.28
		NSBSKDE	*	0.08	0.26
		LRSKDE	*	0.11	0.87
$\mathcal{L}(1)$.2	LR	0.11	0.22	0.99
		NSBS	0.11	0.01	0.01
		NSBSKDE	0.13	0.11	0.02
		LRSKDE	0.12	0.19	0.98
	.4	LR	*	0.14	0.89
		NSBS	*	0.08	0.12
		NSBSKDE	*	0.08	0.03
		LRSKDE	*	0.18	0.86
$\mathcal{U}[-1, 1]$.2	LR	0.19	0.92	1.00
		NSBS	0.11	0.07	0.16
		NSBSKDE	0.11	0.08	0.14
		LRSKDE	0.25	0.99	1.00
	.4	LR	*	0.60	1.00
		NSBS	*	0.09	0.78
		NSBSKDE	*	0.07	0.75
		LRSKDE	*	0.54	1.00

Table 1: Values of the bootstrapped power for LR, NSBS, NSBSKDE and LRSKDE tests when the true density is $\pi^0 f^0(\cdot - u_1^0) + (1 - \pi^0) f^0(\cdot - u_2^0)$, where $u_1^0 = 0$, $u_2^0 - u_1^0 \in \{0, 1, 3\}$, $\pi^0 \in \{0.2, 0.4\}$, and f^0 is one of the zero-symmetric log-concave densities shown in the first column. The nominal level is $\alpha = 0.1$. The sample size is $n = 250$, the number of bootstraps and the number of replications were taken to be $B = 49$ and $R = 500$, respectively. The common value of the power under H_0 is replaced by “*”.

	G	HG	SLC	KDE
$\mathcal{N}(0, 1)$	163 (0.44)	170 (0.30)	170 (0.30)	182 (0.88)
$\mathcal{L}(1)$	223 (0.64)	173 (0.32)	174 (0.28)	154 (0.82)
$\mathcal{U}(-1, 1)$	144 (0.15)	66 (0.11)	55 (0.10)	105 (0.28)

Table 2: Comparison of the four different clustering methods given by the column labels, see text for more details. The reported numbers are the average number of misclassifications out of $n = 500$ samples over $R = 5000$ replications under each of the three log-concave densities in the left column. Here $u_2^0 - u_1^0 = 1$ and $\pi^0 = .2$. The numbers in parentheses are the corresponding standard errors.

two-dimensional log-concave classifier applied to the Breast cancer data of Wisconsin; see [Cule et al. \(2010\)](#) for details. The authors found that the log-concave MLE reduces the percentage of misclassification from 10.36% obtained for the Gaussian estimate to only 8.43% for that particular data set. The posterior probabilities of cluster membership, which can be used as a measurement of uncertainty, can also differ noticeably between the HG method and our SLC method.

6 Data application

In this section, we apply our new estimation approach to two different datasets.

6.1 Old Faithful data

The data to which we first apply our estimation procedure are the times, in minutes, between eruptions of the Old Faithful geyser in Yellowstone National park. There are many forms of the Old Faithful data. As far as we know, the oldest version of the data was collected by S. Weisberg from R. Hutchinson in August 1978. The data we analyze were collected between August 1 and August 15, 1985 continuously, and are from [Azzalini and Bowman \(1990\)](#). The following explanation from [Weisberg \(2005\)](#) motivates interest in the data:

Old Faithful Geyser is an important tourist attraction, with up to several thousand people watching it erupt on pleasant summer days. The park service uses data like these to obtain a prediction equation for the time to the next eruption.

In [Figure 2](#), we have two plots related to the Old Faithful data. The plot on the left depicts a descriptive histogram of the data with around 30 bins (which is too many for optimal estimation) along with the plots of four mixture density estimates. The “SLC” (symmetric log-concave) estimate is the mixture model where u_1^0, u_2^0 and π^0 are estimated using the method of [Hunter et al. \(2007\)](#), and then the components are estimated using our symmetric log-concave estimator. The “HG” (Hunter et al. and Gaussian components) estimate is given by again using the method of [Hunter et al. \(2007\)](#) to find estimates of the mixture parameters whereas the nonparametric components are taken to be Gaussian components (the same Gaussian density for both mixture components). The estimates for u_1^0, u_2^0 , and π^0 given by [Hunter et al. \(2007\)](#) are 55.5, 80.5, and .33, respectively. The “G” (Gaussian) estimate in the plot is based on simply using a Gaussian mixture model with two components with equal variances. Assuming equal variances forces the two components to be identical, which makes the model analogous to the others. In this case, we estimated u_1^0, u_2^0 , and π^0 by the EM algorithm ([Dempster et al., 1977](#)), with estimated values of 55.3, 81.0, and .339. The normal components are slightly more peaked than the log-concave ones, but the overall fit is fairly similar; in large part this is because the locations and weights are very similar. Finally, the ‘KDE’ is a standard kernel density estimator with an optimal bandwidth.

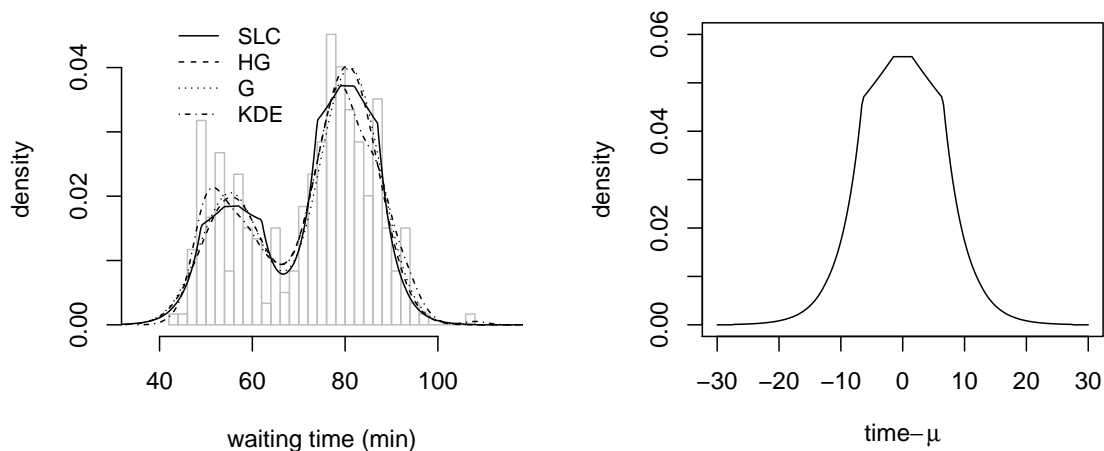


Figure 2: Time between eruptions of Old Faithful Geyser (min). The “SLC” and “HG” estimates both use the method of [Hunter et al. \(2007\)](#) to estimate the mixture parameters u_1^0, u_2^0 , and π^0 . “SLC” then fits with symmetric log-concave components and “HG” fits with Gaussian components. The “G” estimate is the maximum likelihood estimate of a mixture model with two Gaussian components with equal variances. The “KDE” is a standard kernel density estimator with an optimal bandwidth.

The plot on the right is that of the zero-symmetric log-concave component, centered at 0, used in the mixture density. As expected from the known theoretical properties of this estimator, it has a flat interval about the origin, and is the exponential of a concave piecewise linear and zero-symmetric function.

6.2 Height data

We next examine 1766 human height observations. We look at the heights of the population of Campora, a village in the south of Italy. This population is studied by the “Genetic Park of Cilento and Vallo di Dano Project” ([Ciullo, 2009](#)), which is interested in identifying geographically and genetically isolated populations. Such populations are of particular interest because in addition to “genetic homogeneity,” they have a “uniformity of diet, life style and environment.” These homogeneities are valuable in the study of genetic risk factors for complex pathologies such as “hypertension, diabetes, obesity, cancer, and neurodegenerative diseases,” by allowing for a “simplification of the complexity of genetic models” involved, because of the population’s homogeneity ([Ciullo, 2009](#)).

Colonna et al. (2007) provide evidence that this population is indeed genetically isolated. Because of this feature, the distribution of heights of this population is not necessarily the same as that of the global population at large, so estimating its distribution is of interest. Height data are often modeled as mixtures of two components, corresponding to the two sexes.

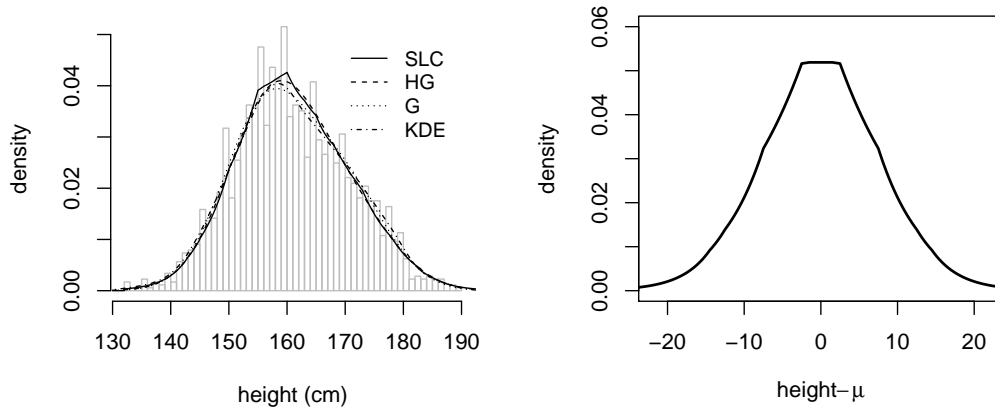


Figure 3: Height data of the population of Campora . The “SLC” and “HG” estimates both use the method of Hunter et al. (2007) to estimate the mixture parameters u_1^0, u_2^0 , and π^0 . “SLC” then fits with symmetric log-concave components and “HG” fits with Gaussian components. The “G” estimate is the maximum likelihood estimate of a mixture model with two Gaussian components with equal variances. The ‘KDE’ is a standard kernel density estimator with an optimal bandwidth.

We present plots related to the height data in Figure 3. The height data do not exhibit multi-modality, but two-component mixtures still fit the data well. The three approaches that we consider fit similarly, but the log-concave components are able to capture a bit more asymmetry near to the mode.

The plot on the right includes the mixture component density (labeled “All”), in black. The data include the sex of each individual, so, using this extra information we can also estimate the true component densities separately: the zero-symmetric log-concave density estimate can be compared to the estimates of the density of the heights for either sex considered alone. Figure 4 shows the plots of the (descriptive) histograms of the heights for men and women and fitted standard kernel density estimators. The assumption of symmetry of the distribution of the heights for each of the genders seems to be reasonable to make. The observed proportion of women is 0.57, whereas the observed medians of the heights for women and men were found to be 156.0 and 168.7, respectively. Using the estimation method of Hunter et al. (2007), we found $\check{\pi}_n = 0.72, \check{u}_{1,n} = 157.5$, and $\check{u}_{2,n} = 170.5$. Here

$\tilde{\pi}_n$ and $\tilde{u}_{1,n}$ correspond to the component for women. The components estimated

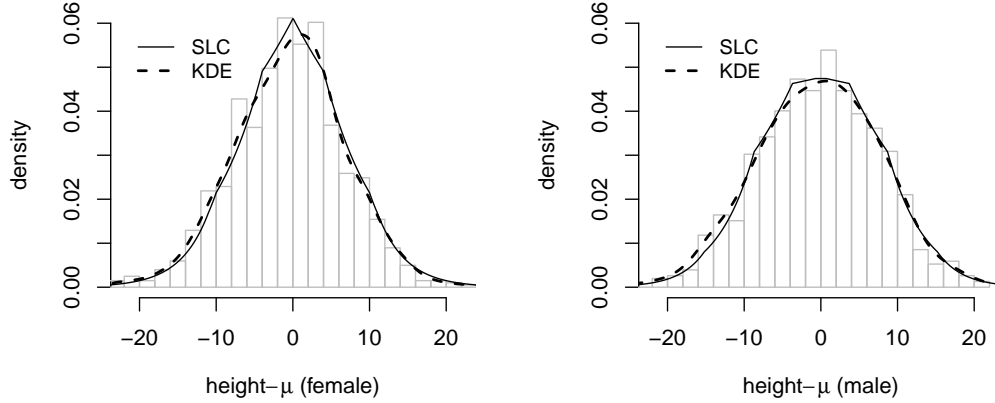


Figure 4: Descriptive histograms of the height data for women (left) and men (right) after centering around the median. The “SLC” is the log-concave MLE of the true density constrained to have mode at 0. The “KDE” is a standard kernel density estimator with an optimal bandwidth.

by using the labels for men and women differ from that using the mixture model without the labels especially towards the center. We believe that this is essentially due to the difference between the estimates of the mixture parameters π^0 , μ_1^0 and μ_2^0 obtained by ignoring or using the information available about the gender. In the latter case, the locations are estimated by the respective medians. It does appear that the distributions of heights of men and women are somewhat different near those centers, with women having a more peaked density and men having a flatter one. Thus, in the mixture model, without using the labels, the component density estimate is somewhere in between the two shapes.

7 Conclusions

The goal in this paper is to make use of the log-concavity constraint to estimate the unknown density component in a semi-parametric location mixture model assuming that this unknown density is symmetric around the origin. The first motivation for choosing this approach is that many densities are log-concave. The second one is to build an estimation procedure that does not depend on a tuning parameter. Our log-concave MLE is computed by maximizing the log-likelihood function after estimating the mixture parameters using the approach of [Hunter et al. \(2007\)](#). The computation is easily implementable using the EM algorithm in combination with an active set algorithm already implemented in the R package **logcondens.mode**.

As already mentioned, our method is not advocated for heavy-tailed densities. In such cases, other shape constraints may be more appropriate, specifically, s -concavity, as studied in [Koenker and Mizera \(2010\)](#) and [Doss and Wellner \(2015\)](#). Unfortunately, the theory of estimators of s -concave densities is less developed than that of log-concave MLEs, which remains a barrier to using s -concavity in our current context.

Finally, [Hunter et al. \(2007\)](#) give sufficient conditions on the mixing probabilities and mixture locations for the model to be 3-identifiable. In this case, the mixture parameters can still be computed using the method of [Hunter et al. \(2007\)](#), and the log-concave MLE can be computed as described in this paper. However, it is not immediate in that case whether the same proof approaches would still yield the same rate of convergence. Recently, [Balabdaoui and Butucea \(2014\)](#) proved that the number of components k , the mixture parameters and the unknown density are identifiable provided that the density is Pólya frequency (of infinite order) such that its expectation is equal to zero. For a precise definition of Pólya frequency functions, we refer to [Schoenberg \(1951\)](#). The obtained identifiability result can be used of course in the case of symmetry but it is certainly not a requirement. Note that imposing the log-concave constraint in this setting is natural since the class of Pólya frequency functions is a subset of the log-concave class as shown by [Schoenberg \(1951\)](#). One may argue that non-parametric classes such as symmetric densities or Pólya frequency functions with expectation equal to zero are not large enough. However, it seems that identifiability is hard to obtain if one allows for large classes.

A Appendix

A.1 Main Proofs

Proof of Proposition 2.1. To show existence, we first start by proving that $\hat{\psi}_n$ is necessarily a piecewise linear function that is flat in the neighborhood of zero. Let $\psi \in \mathcal{SC}$. Also, let $Z_{(1)}$ denote the first order statistic of the transformed data Z_i 's defined in Proposition 2.1. Consider $\bar{\psi}$ the unique concave function such that $\bar{\psi}(Z_i) = \psi(Z_i), i = 1, \dots, 2n$, $\bar{\psi}(0) = \bar{\psi}(Z_{(1)})$, $\bar{\psi}$ is piecewise linear between the points $Z_i, i = 1, \dots, 2n$, $\bar{\psi}(t) = -\infty$ for $t > Z_{(2n)}$ and $\psi(t) = \psi(-t)$ for $t \in \mathbb{R}$. Clearly, $\bar{\psi} \in \mathcal{SC}$ and admits the properties $\tilde{\pi}_n e^{\bar{\psi}(X_i - \tilde{u}_{1,n})} + (1 - \tilde{\pi}_n) e^{\bar{\psi}(X_i - \tilde{u}_{2,n})} = \tilde{\pi}_n e^{\psi(X_i - \tilde{u}_{1,n})} + (1 - \tilde{\pi}_n) e^{\psi(X_i - \tilde{u}_{2,n})}$ for $i = 1, \dots, n$ and $\bar{\psi} \leq \psi$. This implies that $\Phi_n(\bar{\psi}) \geq \Phi_n(\psi)$ and the logarithm of the MLE has to be necessarily piecewise linear between the transformed data points Z_i and $-Z_i, i = 1, \dots, 2n$ with a flat piece in the neighborhood of zero and support $[-Z_{(2n)}, Z_{(2n)}]$. Further, if a maximizer $\hat{f}_n = \exp(\hat{\psi}_n)$ exists then it has to be a density. This follows from the fact that $\hat{\psi}_n + \epsilon \in \mathcal{SC}$ for all $\epsilon \in \mathbb{R}$, and

$$0 = \lim_{\epsilon \rightarrow 0} \frac{\Phi_n(\hat{\psi}_n + \epsilon) - \Phi_n(\hat{\psi}_n)}{\epsilon} = 1 - \int_{\mathbb{R}} \exp(\hat{\psi}_n).$$

Given the results obtained above, a function $\psi \in \mathcal{SC}$ can be identified with the vector

$$\underline{\psi} \equiv \left(\psi(X_i - \check{u}_{1,n}), \psi(X_i - \check{u}_{2,n}) \right)_{i=1}^n = \left(\psi(Z_{2i-1}), \psi(Z_{2i}) \right)_{i=1}^n.$$

Let \mathcal{SC}_n denote the set of such vectors. In this second part of the proof, we will show that the maximization problem admits a solution. It is clear that Φ_n is continuous on \mathcal{SC}_n . In the following, we will show that Φ_n is necessarily maximized on a compact set. To this aim, consider $\underline{\psi}^{(p)}$ to be a maximizing sequence, such that $\int \exp(\psi^{(p)}(t)) dt = 1$, and as $p \rightarrow \infty$

$$\max_{1 \leq i \leq n} |\psi^{(p)}(Z_{(i)})| \rightarrow \infty, \text{ and } |\psi^{(p)}(Z_{(i)})| \rightarrow l_i \in [-\infty, \infty], \text{ for } 1 \leq i \leq n$$

where $Z_{(i)}$ denotes the i -th order statistic of Z_1, \dots, Z_{2n} .

For $1 \leq i \leq n$ write $\psi_i^{(p)} := \psi^{(p)}(Z_{(i)})$. Suppose that for all $i \in \{1, \dots, n\}$

$$\lim_{p \rightarrow \infty} \psi_i^{(p)} < \infty.$$

This implies that there exists $j \in \{1, \dots, n\}$ such that $\lim_{p \rightarrow \infty} \psi_j^{(p)} = -\infty$. If there exists $i \in \{1, \dots, n\}$ such that $Z_{(j)} = \max(|X_i - \check{u}_{1,n}|, |X_i - \check{u}_{2,n}|)$ then

$$\begin{aligned} \lim_{p \rightarrow \infty} \log \left(\check{\pi}_n \exp(\psi^{(p)}(|X_i - \check{u}_{1,n}|)) + (1 - \check{\pi}_n) \exp(\psi^{(p)}(|X_i - \check{u}_{2,n}|)) \right) &\leq \lim_{p \rightarrow \infty} \psi_j^{(p)} \\ &= -\infty \end{aligned}$$

implying that $\lim_{p \rightarrow \infty} \Phi_n(\psi^{(p)}) = -\infty$ which is in contradiction with the definition of $\psi^{(p)}$.

Let us assume now that for any j such that $\lim_{p \rightarrow \infty} \psi_j^{(p)} = -\infty$, there exists $i \in \{1, \dots, n\}$ such that

$$Z_{(j)} = \min(|X_i - \check{u}_{1,n}|, |X_i - \check{u}_{2,n}|) \text{ and } \lim_{n \rightarrow \infty} \psi^{(p)}(\max(|X_i - \check{u}_{1,n}|, |X_i - \check{u}_{2,n}|)) > -\infty. \quad (13)$$

Suppose that $j = 2n$ is the only integer for which this divergence is occurring. Let $i \in \{1, \dots, n\}$ such that we have $Z_{(2n)} = \min(|X_i - \check{u}_{1,n}|, |X_i - \check{u}_{2,n}|)$. Without loss of generality, assume that $|X_i - \check{u}_{1,n}| < |X_i - \check{u}_{2,n}|$. Put

$$a^{(p)} = (1 - \check{\pi}_n) \exp \left(\psi^{(p)}(|X_i - \check{u}_{2,n}|) \right), \text{ and } b^{(p)} = \psi_{2n-1}^{(p)}$$

and call a and b their respective limits as $p \rightarrow \infty$. Note that both a and b are in \mathbb{R} by assumption. Consider now the function defined by

$$h(x) = \log(\check{\pi}_n e^x + a) - 2(Z_{(2n)} - Z_{(2n-1)}) \frac{\exp(b) - \exp(x)}{b - x}$$

on $(-\infty, 0]$. Put $\Delta_{2n-1} = Z_{(2n)} - Z_{(2n-1)}$. We have that

$$h'(x) = \frac{\check{\pi}_n \exp(x)}{\check{\pi}_n \exp(x) + a} - \frac{2\Delta_{2n-1}}{b-x} \left(\frac{\exp(b) - \exp(x)}{b-x} - \exp(x) \right)$$

which has the same sign as

$$\begin{aligned} & \check{\pi}_n - 2\Delta_{2n-1} \exp(b) \left(\check{\pi}_n \exp(x) + a \right) \frac{1}{b-x} \left(\frac{\exp(-x) - \exp(-b)}{b-x} - \exp(-b) \right) \\ &= \check{\pi}_n - 2\Delta_{2n-1} \exp(b) \left(\check{\pi}_n \exp(x) + a \right) \frac{\exp(-x) - \exp(-b)}{(b-x)^2} \left(1 - \frac{(b-x) \exp(-b)}{\exp(-x) - \exp(-b)} \right) \\ &\rightarrow -\infty, \text{ as } x \rightarrow -\infty. \end{aligned}$$

This implies that as $p \rightarrow \infty$

$$\log \left(\check{\pi}_n \exp(\psi^{(p)}(|X_i - \check{u}_{1,n}|)) + (1 - \check{\pi}_n) \exp(\psi^{(p)}(|X_i - \check{u}_{2,n}|)) \right) - \int_{Z_{(2n-1)}}^{Z_{(2n)}} \exp(\psi^{(p)}(t)) dt$$

must be decreasing, and hence bigger values of the log-likelihood are to be obtained if $\psi^{(p)}(Z_{(2n)})$ were not divergent. This implies that $\psi_{2n}^{(p)}$ cannot diverge to $-\infty$. The same reasoning can be applied if (13) is satisfied by other integers $k \in \{1, \dots, 2n-1\}$, and let j be the smallest one. Let i be such that $Z_{(j)} = \min(|X_i - \check{u}_{1,n}|, |X_i - \check{u}_{2,n}|)$. Note that $j > 1$. Also, note that by definition of j $\lim_{p \rightarrow \infty} \psi^{(p)}(Z_{(j-1)}) > -\infty$. Thus, we obtain the same conclusion as before, that is, the term

$$\log \left(\check{\pi}_n \exp(\psi^{(p)}(|X_i - \check{u}_{1,n}|)) + (1 - \check{\pi}_n) \exp(\psi^{(p)}(|X_i - \check{u}_{2,n}|)) \right) - \int_{Z_{(j-1)}}^{Z_{(j)}} \exp(\psi^{(p)}(t)) dt$$

can be increased (and hence the log-likelihood) if $\psi_j^{(p)}$ were not diverging to $-\infty$. By a recursive reasoning, we obtain the same conclusion about the remaining integers.

Now suppose that $\lim_{p \rightarrow \infty} \max_{1 \leq i \leq n} \psi^{(p)}(Z_{(i)}) = \infty$. Since $\psi^{(p)}$ is decreasing on $[0, \infty)$, $\max_{1 \leq i \leq n} \psi^{(p)}(Z_{(i)}) = \psi^{(p)}(Z_{(1)}) \rightarrow \infty$ as $p \rightarrow \infty$. For $x \in [Z_{(1)}, Z_{(2)}]$, we have

$$\begin{aligned} 1 &\geq \int_{Z_{(1)}}^{Z_{(2)}} \exp(\psi_1^{(p)}) \exp \left(\frac{\psi_2^{(p)} - \psi_1^{(p)}}{Z_{(2)} - Z_{(1)}} (t - Z_{(1)}) \right) dt \\ &= \exp(\psi_1^{(p)})(Z_{(2)} - Z_{(1)}) \frac{1 - \exp(\psi_1^{(p)} - \psi_2^{(p)})}{\psi_1^{(p)} - \psi_2^{(p)}} \\ &\geq \exp(\psi_1^{(p)})(Z_{(2)} - Z_{(1)}) \frac{1}{\psi_1^{(p)} - \psi_2^{(p)} + 1}, \end{aligned}$$

where the last inequality follows from the fact that $\frac{1 - \exp(-t)}{t} \geq \frac{1}{t+1}$ on $[0, \infty)$. This implies that

$$\psi_1^{(p)} - \psi_2^{(p)} \geq \exp(\psi_1^{(p)})(Z_{(2)} - Z_{(1)}) - 1.$$

Thus

$$\psi_1^{(p)} + \psi_2^{(p)} = 2\psi_1^{(p)} + \psi_2^{(p)} - \psi_1^{(p)} \leq 2\psi_1^{(p)} - \exp(\psi_1^{(p)})(Z_{(2)} - Z_{(1)}) + 1 \rightarrow -\infty$$

as $p \rightarrow \infty$, and hence $\psi_2^{(p)} \leq (\psi_1^{(p)} + \psi_2^{(p)})/2 \rightarrow -\infty$. Since a maximizer of Φ_n is necessarily constant on $[-\tau_1, \tau_1]$ with τ_1 the first kink point, here equal to $Z_{(1)}$, we can write

$$1 = 2 \exp(\psi_1^{(p)})Z_{(1)} + \sum_{j=2}^n \int_{Z_{(j)}}^{Z_{(j+1)}} \exp(\psi^{(p)}(t))dt \geq 2 \exp(\psi_1^{(p)})Z_{(1)}$$

But this implies using the results obtained above that $\exp(\psi_1^{(p)}) \leq 1/(2Z_{(1)})$, which contradicts the assumption that $\psi_1^{(p)} \rightarrow \infty$. \square

Proof of Proposition 2.2. This follows from arguments similar to those used in the proof of Theorem 2.2 of [Dümbgen and Rufibach \(2009\)](#). \square

Proof of Proposition 2.3: Using symmetry of \hat{f}_n and Δ , and the definition of $\hat{\mathbb{F}}_n$, the inequality in (7) can be re-written as

$$\int_0^\infty \Delta(x) d\hat{\mathbb{F}}_n(x) \leq \int_0^\infty \Delta(x) \hat{f}_n^+(x) dx, \quad (14)$$

with equality for Δ satisfying $\hat{\psi}_n^+ \pm \epsilon \Delta \in \mathcal{SC}$ for $\epsilon > 0$ small enough. For $z \in [0, \infty)$, consider the concave perturbation function defined on $[0, \infty)$ as

$$\begin{aligned} \Delta_z(x) &= -(x - z)_+ \\ &= \begin{cases} z - x & \text{if } x \geq z \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

Using Fubini's theorem, the inequality in (14) yields

$$Z_{(2n)} - z - \int_z^{Z_{(2n)}} \hat{\mathbb{F}}_n(x) dx \geq Z_{(2n)} - z - \int_z^{Z_{(2n)}} \hat{F}_n^+(x) dx.$$

This yields

$$\int_0^z \hat{F}_n^+(x) dx \leq \int_0^z \hat{\mathbb{F}}_n(x) dx + \int_0^{Z_{(2n)}} \hat{F}_n^+(x) dx - \int_0^{Z_{(2n)}} \hat{\mathbb{F}}_n(x) dx.$$

Since for $z = 0$, the inequality in (14) becomes an equality (Δ_0 is a straight line, and hence satisfies $\hat{\psi}_n^+ \pm \epsilon \Delta_0 \in \mathcal{SC}$ for $\epsilon > 0$ small enough), we have that

$$\int_0^{Z_{(2n)}} \hat{\mathbb{F}}_n(x) dx = \int_0^{Z_{(2n)}} \hat{F}_n^+(x) dx$$

from which follows the inequality part in (9). The equality part follows from noting that when z is a knot of $\hat{\psi}_n^+ = \log(\hat{f}_n^+)$ we have $\hat{\psi}_n^+ \pm \epsilon \Delta_z \in \mathcal{SC}$ for $\epsilon > 0$ small enough. \square

Next we state and prove propositions and lemmas needed for the proofs of the theorems in Section 3, and then prove those theorems. We start by stating several propositions of the Glivenko-Cantelli type. To begin, let \mathcal{C} be the class of convex and compact subsets (intervals) in \mathbb{R} . The following theorem of Bhattacharya and Ranga Rao (1976) will be an important component of the proof of Proposition A.1

Theorem A.1 (Theorem 1.11, page 22, Bhattacharya and Ranga Rao (1976))

Let P be a probability measure on \mathbb{R} . Then, as $n \rightarrow \infty$

$$\sup_{C \in \mathcal{C}} |\mathbb{P}_n(C) - P(C)| \rightarrow 0$$

almost surely.

Proposition A.1 Fix $\alpha \in (0, 1)$. Let X_1, \dots, X_n be n independent observations from the density $g^0 = \tilde{\pi}_n^0 f^0(\cdot - u_1^0) + (1 - \tilde{\pi}_n^0) f^0(\cdot - u_2^0)$, and $\tilde{u}_{1,n}$ and $\tilde{u}_{2,n}$ be any consistent estimators of u_1^0 and u_2^0 . Write $\bar{u}_n = (\tilde{u}_{1,n} + \tilde{u}_{2,n})/2$, $\bar{u}^0 = (u_1^0 + u_2^0)/2$ and let $\bar{N} = \sum_{i=1}^n 1_{X_i \leq \bar{u}}$. Then, for any $\epsilon > 0$, there exists n_0 such that

$$P \left(\min \left\{ |C| : C \in \mathcal{C} \text{ such that } \bar{N}^{-1} \sum_{i: X_i \leq \bar{u}_n} 1_{\{X_i \in C\}} \geq \alpha \right\} \geq \frac{\alpha G(\bar{u}^0)}{2 \|g^0\|_\infty} \right) \geq 1 - \epsilon$$

for all $n \geq n_0$.

Proof: Let \mathcal{C} denote the class of all convex and compact subsets in \mathbb{R} . Fix $\epsilon > 0$. For $C \in \mathcal{C}$ we have that

$$\bar{N}^{-1} \sum_{i: X_i \leq \bar{u}} 1_{\{X_i \in C\}} = \frac{\mathbb{P}_n(C \cap]-\infty, \bar{u}])}{\mathbb{G}_n(\bar{u})} = \frac{\mathbb{P}_n(C_{\bar{u}})}{\mathbb{G}_n(\bar{u})}.$$

where $C_{\bar{u}} = C \cap]-\infty, \bar{u}]$ is also compact and convex. To simplify notation, we write C for $C_{\bar{u}_n}$, and g , G and P the true density, distribution function and associated measure respectively. Then

$$\begin{aligned} \frac{\mathbb{P}_n(C)}{\mathbb{G}_n(\bar{u}_n)} - \frac{P(C)}{G(\bar{u}^0)} &= \frac{[\mathbb{P}_n(C) - P(C)]G(\bar{u}_n) - P(C)[\mathbb{G}_n(\bar{u}_n) - G(\bar{u}_n)]}{\mathbb{G}_n(\bar{u}_n)G(\bar{u}_n)} \\ &\quad - \frac{P(C)[G(\bar{u}_n) - G(\bar{u}^0)]}{G(\bar{u}_n)G(\bar{u}^0)}. \end{aligned}$$

Consistency of $\tilde{u}_{1,n}$ and $\tilde{u}_{2,n}$, the fact that the class $\{[-\infty, x] : x \in \mathbb{R}\}$ is Glivenko-Cantelli (see Example 2.5.4 of [van der Vaart and Wellner \(1996\)](#)) and continuity of g imply that the event

$$\left\{ G(\bar{u}_n) \geq \frac{G(\bar{u}^0)}{2}, \mathbb{G}_n(\bar{u}_n) \geq \frac{G(\bar{u}^0)}{2}, \max_{t \in [\bar{u}^0 \wedge \bar{u}_n, \bar{u}^0 \vee \bar{u}_n]} g(t) \leq 2g(\bar{u}^0) \right\}$$

occurs with probability converging to one. On the other hand, we have that

$$\begin{aligned} \left| \frac{\mathbb{P}_n(C)}{\mathbb{G}_n(\bar{u}_n)} - \frac{P(C)}{G(\bar{u}^0)} \right| &\leq \frac{\sup_{C \in \mathcal{C}} |\mathbb{P}_n(C) - P(C)| + \sup_{x \in \mathbb{R}} |\mathbb{G}_n(x) - G(x)|}{\mathbb{G}_n(\bar{u}_n)G(\bar{u}_n)} \\ &\quad + \frac{|G(\bar{u}_n) - G(\bar{u}^0)|}{G(\bar{u}_n)G(\bar{u}^0)} \\ &\leq \frac{\sup_{C \in \mathcal{C}} |\mathbb{P}_n(C) - P(C)| + \sup_{x \in \mathbb{R}} |\mathbb{G}_n(x) - G(x)|}{\mathbb{G}_n(\bar{u}_n)G(\bar{u}_n)} \\ &\quad + \frac{g(\xi)|\bar{u}_n - \bar{u}^0|}{G(\bar{u}_n)G(\bar{u}^0)}, \text{ for some } \xi \in [\bar{u}^0 \wedge \bar{u}_n, \bar{u}^0 \vee \bar{u}_n] \\ &\leq D \left(\sup_{C \in \mathcal{C}} |\mathbb{P}_n(C) - P(C)| + \sup_{x \in \mathbb{R}} |\mathbb{G}_n(x) - G(x)| + |\bar{u}_n - \bar{u}^0| \right) \end{aligned}$$

probability increasing to one, where $D = 4G(\bar{u}^0)^{-2} \max(2g(\bar{u}^0), 1)$. Using Theorem A.1, and again consistency of \bar{u}_n , we conclude that

$$\sup_{C \in \mathcal{C}} \left| \frac{\mathbb{P}_n(C)}{\mathbb{G}_n(\bar{u}_n)} - \frac{P(C)}{G(\bar{u}^0)} \right| \rightarrow 0 \text{ in probability as } n \rightarrow \infty.$$

In particular, this implies that for any fixed $\epsilon > 0$ there exists n_0 such that

$$P \left(\frac{\mathbb{P}_n(C)}{\mathbb{G}_n(\bar{u}_n)} \leq \frac{P(C)}{G(\bar{u}^0)} + \frac{\alpha}{2} \text{ for all } C \in \mathcal{C} \right) \geq 1 - \epsilon.$$

Let us denote by $\mathcal{E}_{n,\alpha/2}$ the event $\left\{ \frac{\mathbb{P}_n(C)}{\mathbb{G}_n(\bar{u}_n)} \leq \frac{P(C)}{G(\bar{u}^0)} + \frac{\alpha}{2} \text{ for all } C \in \mathcal{C} \right\}$. Also, consider the subclass

$$\mathcal{C}_{\alpha/2} = \left\{ C \in \mathcal{C} : \frac{P(C)}{G(\bar{u}^0)} \geq \frac{\alpha}{2} \right\}.$$

Clearly, for any $C \in \mathcal{C}_{\alpha/2}$ we have that

$$|C| \geq \frac{\alpha G(\bar{u}^0)}{2 \|g\|_\infty}.$$

On the other hand,

$$\begin{aligned} &\left\{ C \in \mathcal{C} : \bar{N}^{-1} \sum_{i: X_i \leq \bar{u}_n} 1_{\{X_i \in C\}} \geq \alpha \right\} = \left\{ C \in \mathcal{C} : \frac{\mathbb{P}_n(C)}{\mathbb{G}_n(\bar{u}_n)} \geq \alpha \text{ for } C \in \mathcal{C} \right\} \\ &= \left\{ \left\{ C \in \mathcal{C} : \frac{\mathbb{P}_n(C)}{\mathbb{G}_n(\bar{u}_n)} \geq \alpha \right\} \cap \mathcal{E}_{n,\alpha/2} \right\} \cup \left\{ \left\{ C \in \mathcal{C} : \frac{\mathbb{P}_n(C)}{\mathbb{G}_n(\bar{u}_n)} \geq \alpha \right\} \cap \mathcal{E}_{n,\alpha/2}^c \right\} \\ &\subseteq \mathcal{C}_{\alpha/2} \cup \left\{ \left\{ C \in \mathcal{C} : \frac{\mathbb{P}_n(C)}{\mathbb{G}_n(\bar{u}_n)} \geq \alpha \right\} \cap \mathcal{E}_{n,\alpha/2}^c \right\} \end{aligned}$$

implying that for $n \geq n_0$

$$P \left(\min \{ |C| : C \in \mathcal{C} \text{ such that } \bar{N}^{-1} \sum_{i: X_i \leq \bar{u}_n} 1_{\{X_i \in C\}} \geq \alpha \} < \frac{\alpha G(\bar{u}^0)}{2 \|g^0\|_\infty} \right) \leq \epsilon$$

completing the proof of the Proposition. \square

Lemma A.1 *Let $\psi \in \mathcal{SC}$, $a < b \in \mathbb{R}$ and $\lambda \in [0, 1]$. If $p(t) = \lambda \exp(\psi(t-a)) + (1-\lambda) \exp(\psi(t-b))$, then*

$$\int_{\mathbb{R}} |\log(p(t))| p(t) dt < \infty.$$

Proof: Note that for $t \in \mathbb{R}$ we have that

$$|t-a| \leq |t-b| \iff t \leq (a+b)/2.$$

Using the fact that the function $t \mapsto |\log(t)|t$ is increasing on $(0, e^{-1}]$ and that ψ is decreasing on $[0, \infty)$, we have for A taken to be larger than $(a+b)/2$,

$$\begin{aligned} \int_A^\infty |\log(p(t))| p(t) dt &\leq \int_A^\infty |\psi(t-a)| \exp(\psi(t-a)) dt \\ &< \int_\infty^\infty \end{aligned}$$

where the second inequality follows e.g. from Lemma 2 in [Schuhmacher and Dümbgen \(2010\)](#). By symmetry, the integral on $(-\infty, -A]$ is also finite. \square

In the following, we introduce a condition that will be useful in proving consistency of our estimator.

The condition (C): We will say that an (ordered) pair $(f, h) = (e^\psi, e^\phi)$ of log-concave densities satisfy the condition (C) if for all $d \in \mathbb{R}$,

$$\int_0^\infty |\phi(t+d)| f(t) dt < \infty.$$

Note that h must have support $(-\infty, \infty)$.

Lemma A.2 *Let ψ_0 and ϕ be elements of \mathcal{SC} . Let $a_0 < b_0$ be real numbers and $\lambda_0 \in [0, 1]$. Let (a_k) , (b_k) , and (λ_k) , for $k > 0$, be sequences of real numbers converging to a_0, b_0 , and λ_0 , respectively. Let*

$$\begin{aligned} p_k(\cdot) &= \lambda_k \exp(\psi_0(\cdot - a_k)) + (1 - \lambda_k) \exp(\psi_0(\cdot - b_k)), \\ p_{k,\phi}(\cdot) &= p_k(\cdot) \star e^{\phi(\cdot)} = \lambda_k \exp(\psi_0(\cdot - a_k)) \star e^{\phi(\cdot)} + (1 - \lambda_k) \exp(\psi_0(\cdot - b_k)) \star e^{\phi(\cdot)}, \end{aligned}$$

for $k = 0, 1, \dots$. Let e^{ψ_0} and e^ϕ satisfy condition (C). Then

$$\int_{\mathbb{R}} \log(p_{k,\phi}(t)) p_0(t) dt \longrightarrow \int_{\mathbb{R}} \log(p_{0,\phi}(t)) p_0(t) dt \quad (15)$$

as $k \rightarrow \infty$.

Proof: It is enough to show that for k large, $|\log p_{k,\phi}|$ is bounded by a p_0 -integrable function to apply the Lebesgue Dominated Convergence Theorem. Since $p_{k,\phi}$ is uniformly bounded above, to upper bound $|\log p_{k,\phi}|$ we need to lower bound $p_{k,\phi}$. Now,

$$p_{k,\phi}(x) = \int_{\mathbb{R}} p_k(y) e^{\phi(x-y)} dy \geq \int_{\alpha}^{\beta} \left(\min_{z \in [\alpha, \beta]} p_k(z) \right) e^{\phi(x-y)} dy \quad (16)$$

for any α and β . Taking α and β to be in the interior of the support of p_0 , we have that, for all k large enough, $\min_{z \in [\alpha, \beta]} p_k(z) \geq c$ for some $c > 0$. Now because ϕ is concave, if $x \geq A$ where A is large enough, ϕ is decreasing on $[x - \beta, x - \alpha]$, i.e. $\phi(x - y)$ is increasing in y for $y \in [\alpha, \beta]$, so that then $\phi(x - y) \geq \phi(x - \alpha)$. Taking A large enough that $\phi(x - \alpha) < 0$, so $|\phi(x - y)| \leq |\phi(x - \alpha)|$, we have that, for k large, and $x > A$, (16) is bounded below by

$$c \int_{\alpha}^{\beta} e^{\phi(x-\alpha)} dy = (\beta - \alpha) c e^{\phi(x-\alpha)}.$$

Recall that if $d_1 < d_2$, then

$$|t - d_1| \leq |t - d_2| \iff t \leq (d_1 + d_2)/2.$$

Thus, taking $A > (a + b)/2$,

$$\int_A^{\infty} |\phi(x - \alpha)| p_0(x) dx \leq \int_A^{\infty} |\phi(x - \alpha)| e^{\psi_0(x-b_0)} dx = \int_{A+b}^{\infty} |\phi(x - \alpha + b)| e^{\psi_0(x)} dx,$$

and this is less than infinity, by the condition (C). Thus, for $x \geq A$ and k large,

$$|\log p_{k,\phi}(x)| \leq |\log((\beta - \alpha)c)| + |\phi(x - \alpha)|,$$

so we have shown that $|\log p_{k,\phi}|$ is bounded by a function p_0 -integrable on $[A, \infty)$. A similar argument shows it is bounded by a function integrable on $(-\infty, -A]$. Finally, since $p_{0,\phi}$ has support $(-\infty, \infty)$ (since e^{ϕ} has support $(-\infty, \infty)$ by the condition (C)),

$$\sup_{x \in [-A, A]} |\log p_{k,\phi}(x)| \leq \sup_{x \in [-A-\eta, A+\eta]} |\log p_0(x)| < \infty,$$

for all k large enough that $\max(|a_0 - a_k|, |b_0 - b_k|) < \eta$ for some $\eta > 0$. Thus for k large we have shown that $|\log p_{k,\phi}|$ is bounded by a p_0 -integrable function on \mathbb{R} , and $p_{k,\phi}(t) \rightarrow p_{0,\phi}(t)$ for all $t \in \mathbb{R}$, so (15) holds. \square

Note that for any log-concave density $f = e^{\psi}$, if ϕ is a polynomial function, then f and e^{ϕ} satisfy condition (C), since log-concave densities have finite moments of all orders. In particular, we may take e^{ϕ} to be any normal density, which we will do below in the proof of consistency. Theorem 3.1 will rely on the following two Glivenko-Cantelli propositions. As is standard, we say that a class of functions \mathcal{F} is a *G-Glivenko-Cantelli class* if $\sup_{f \in \mathcal{F}} |\mathbb{G}_n(f) - G(f)| \rightarrow 0$ as $n \rightarrow \infty$ almost surely,

where \mathbb{G}_n is the empirical probability measure based on i.i.d. draws from G , and we write $\mu(f) = \int f d\mu$ for a function f and a measure μ (as in e.g. [van der Vaart and Wellner \(1996\)](#) or [van de Geer \(2000\)](#)). (This is sometimes known as being *strong* Glivenko-Cantelli because the convergence is almost sure, rather than in probability ([van der Vaart and Wellner, 2000](#)).)

Proposition A.2 *Let $M > 0$ and \mathcal{LC}_M be the class of log-concave functions bounded by M . For a fixed $b > 0$, consider the class of functions*

$$\mathcal{F}_b = \left\{ f : f = \log(\lambda h_1 + (1 - \lambda)h_2 + b), \ h_1, h_2 \in \mathcal{LC}_M, \ \lambda \in [0, 1] \right\}.$$

Then \mathcal{F}_b is a Glivenko-Cantelli class for any probability measure P .

Proof: Since all $f \in \mathcal{LC}_M$ are unimodal, $\mathcal{LC}_M \subset \mathcal{M}_M^+ - \mathcal{M}_M^+$ where $\mathcal{M}_M^+ := \{f : 0 \leq f \leq M, f \text{ is nondecreasing}\}$. By Theorem 2.7.5 and Theorem 2.4.1 in [van der Vaart and Wellner \(1996\)](#), \mathcal{M}_M^+ is a (strong) P -Glivenko-Cantelli class for any P . By the Glivenko-Cantelli preservation Theorem 3 of [van der Vaart and Wellner \(2000\)](#), $\varphi(\mathcal{M}_M^+, \mathcal{M}_M^+)$ is P -Glivenko-Cantelli, with $\varphi(x, y) := x - y$ here, so \mathcal{LC}_M also is P -Glivenko-Cantelli, since it has an integrable envelope by construction. Now letting $\varphi(h_1, h_2, \lambda) := \log(\lambda h_1 + (1 - \lambda)h_2 + b)$, another application of Theorem 3 of [van der Vaart and Wellner \(2000\)](#) yields that $\mathcal{F}_b = \varphi(\mathcal{LC}_M, \mathcal{LC}_M, \{\lambda \mathbb{1}_{(-\infty, \infty)}(x) : \lambda \in [0, 1]\})$ is also Glivenko-Cantelli since each of the three argument classes are, as long as \mathcal{F}_b has a P -integrable envelope. But this is immediate since $\log(b) \leq f \leq \log(M + b)$ for all $f \in \mathcal{F}_b$. \square

We will also need a Glivenko-Cantelli result for the case $b = 0$. In order for such a result to hold, we need to restrict attention to just logs of mixtures of a single density; despite the unboundedness of the functions involved, this will allow us to have an integrable envelope, at least for the true P_0 .

Proposition A.3 *Fix $\tilde{f}_0 \in \mathcal{SLC}_1$, $\tilde{\pi}_n \in [0, 1]$, $a < b$ and let $p_0(x) = \tilde{\pi}_n \tilde{f}_0(x - a) + (1 - \tilde{\pi}_n) \tilde{f}_0(x - b)$, $x \in \mathbb{R}$, and P_0 the corresponding probability measure. Let $q \in \mathcal{SLC}_1$ be such that (\tilde{f}_0, q) satisfy the Condition (C). Then the class of functions*

$$\mathcal{H} = \left\{ \log(\lambda q(x - \alpha) + (1 - \lambda)q(x - \beta)) : \right. \\ \left. \alpha \in [a - r, a + r], \beta \in [b - r, b + r], \lambda \in [0, 1] \right\}$$

for any fixed $r > 0$ is P_0 -Glivenko-Cantelli.

Proof: Without loss of generality, we take $r = 1$ in this proof. For any $c \in \mathbb{R}$ let $\mathcal{P}_c := \{q(\cdot - z) : z \in [c - 1, c + 1]\}$. Then the classes \mathcal{P}_a and \mathcal{P}_b are both contained in \mathcal{LC}_M with $M = \sup_x q(x)$, and in the proof of Proposition A.2 we showed that \mathcal{LC}_M is P -Glivenko-Cantelli for any P . Let $T(h_1, h_2, \lambda) := \log(\lambda h_1 + (1 - \lambda)h_2)$

and let $I_{0,1} := \{\lambda \mathbb{1}_{(-\infty, \infty)}(x) : \lambda \in [0, 1]\}$. Then by Theorem 3 of [van der Vaart and Wellner \(2000\)](#), the class $T(\mathcal{P}_a, \mathcal{P}_b, I_{0,1})$ is P_0 -Glivenko-Cantelli as long as it has a P_0 -integrable envelope, which is what we will show now.

Let $c_+ \in \mathbb{R}$ be such that $\log q(x - (b + 1)) < 0$ if $x > c_+$. Then for any $h \in \mathcal{H}$, $h(x) \leq 0$ for $x > c_+$ by the unimodality of q . Note $q(x - \alpha) \geq q(x - (a - 1))$ for $x \geq a + 1 \geq \alpha \geq a - 1$ and $q(x - \beta) \geq q(x - (b - 1))$ if $x \geq b + 1 \geq \beta \geq b - 1$. Thus, for $x > (b + 1) \vee c_+$, $h \in \mathcal{H}$ satisfies

$$\begin{aligned} 0 \geq h(x) &\geq \log(\lambda q(x - (a - 1)) + (1 - \lambda)q(x - (b - 1))) \\ &\geq \log q(x - (a - 1)), \end{aligned} \tag{17}$$

since $b > a$. A similar argument shows for $x < (a - 1) \wedge c_-$ for some $c_- \in \mathbb{R}$, that $0 \geq h(x) \geq \log q(x - (b + 1))$. By the assumption that Condition (C) holds for (\tilde{f}_0, q) , it is immediate that $\log q(\cdot - (a - 1))$ and $\log q(\cdot - (b + 1))$ are p_0 -integrable. All $h \in \mathcal{H}$ are uniformly bounded on $[(a - 1) \wedge c_-, (b + 1) \vee c_+]$, since the support of $p_{0,\varphi}$ is \mathbb{R} , since $p_{0,\varphi}$ is unimodal, and since the range of α and β in the definition of \mathcal{H} is bounded. Thus \mathcal{H} has a P_0 -integrable envelope and so we are done. \square

Theorem A.2 *Let \hat{g}_n denote again the MLE of the mixed density. Then, for a given $\epsilon > 0$ there exist an integer $n_0 > 0$ and $D > 0$ depending on ϵ such that*

$$P(\|\hat{g}_n\|_\infty \leq D) \geq 1 - \epsilon.$$

for all $n \geq n_0$.

Proof: Fix $\epsilon > 0$. Recall that if $a < b$, then for $t \in \mathbb{R}$

$$|t - a| \leq |t - b| \iff t \leq (a + b)/2.$$

Write $\bar{u} = (\check{u}_{1,n} + \check{u}_{2,n})/2$. Using the symmetry of $\hat{\psi}_n$ and the fact that it is decreasing on $[0, \infty)$, it follows that

$$\Phi_n(\log(\hat{g}_n)) \leq \frac{1}{n} \sum_{i: X_i \leq \bar{u}} \hat{\psi}_n(X_i - \check{u}_{1,n}) + \frac{1}{n} \sum_{i: X_i > \bar{u}} \hat{\psi}_n(X_i - \check{u}_{2,n}) - 1. \tag{18}$$

Without loss of generality, we can assume that $\{X_1, X_2, \dots, X_{\bar{N}}\} = \{X_i : X_i \leq \bar{u}\}$ and that $\hat{\psi}_n(X_1 - \check{u}_{1,n}) < \dots < \hat{\psi}_n(X_{\bar{N}} - \check{u}_{1,n})$, and that

$$\max_{i: X_i > \bar{u}} \hat{\psi}_n(X_i - \check{u}_{2,n}) \leq \max_{i: X_i \leq \bar{u}} \hat{\psi}_n(X_i - \check{u}_{1,n})$$

otherwise the same reasoning above has to be applied to the second term of the right side of the inequality in (18). In the following, let

$$M_n = \hat{\psi}_n(X_n - \check{u}_{1,n})$$

the largest value taken by $\hat{\psi}_n(X_i - \check{u}_{1,n})$ and $\hat{\psi}_n(X_i - \check{u}_{2,n})$ for $i = 1, \dots, n$.

Following the notation of [Schuhmacher and Dümbgen \(2010\)](#) in the proof of their Lemma 4, let $k = \lfloor \bar{N}\alpha \rfloor$ for some fixed $\alpha \in (0, 1]$, \mathcal{C}_k the convex hull of the set $\{X_i - \check{u}_{1,n}, i \geq k\}$ and $m = \hat{\psi}_n(X_k - \check{u}_{1,n})$. Also, let

$$\eta = \frac{\alpha G(\bar{u}^0)}{2\|g^0\|_\infty}.$$

In the sequel, $\mathcal{C}_k + \check{u}_{1,n} = \{t + \check{u}_{1,n} : t \in \mathcal{C}_k\}$. Note that $\mathcal{C}_k + \check{u}_{1,n}$ is convex and compact such that

$$\sum_{i: X_i \leq \bar{u}} 1_{\{X_i \in \mathcal{C}_k + \check{u}_{1,n}\}} = \sum_{i=1}^{\bar{N}} 1_{\{X_i \in \mathcal{C}_k + \check{u}_{1,n}\}} = \bar{N} - k + 1 \geq \bar{N}(1 - \alpha).$$

By taking $1 - \alpha$ in place of α in Proposition [A.1](#), we can find an integer $n_0 > 0$ such that for $n \geq n_0$.

$$P(|\mathcal{C}_k| \geq \eta) \geq 1 - \epsilon/5. \quad (19)$$

Consider the event

$$\left\{M_n > B_0 := \max(\log(1/\eta) + 1, 0)\right\}. \quad (20)$$

Using the same argument as in [Schuhmacher and Dümbgen \(2010\)](#), we can write

$$\begin{aligned} 1 = \int_{\mathbb{R}} e^{\hat{\psi}_n(t)} dt &\geq \int_{\mathcal{C}_k} e^{\hat{\psi}_n(t)} dt \\ &\geq e^m |\mathcal{C}_k| \end{aligned} \quad (21)$$

using monotonicity of $\hat{\psi}_n$ on the convex set \mathcal{C}_k . By [\(20\)](#). Hence,

$$\left\{M_n > B_0, |\mathcal{C}_k| > \eta\right\} \subset \{M_n > m + 1\}. \quad (22)$$

Indeed, this follows from the fact that

$$\begin{aligned} e^{M_n-1} &> 1/\eta, \quad \text{by (20)} \\ &> 1/|\mathcal{C}_k| \\ &\geq e^m \text{ by (21)}. \end{aligned}$$

Also, on the event on the right side of [\(22\)](#), we have for t such that $t \in \mathcal{C}_k + \check{u}_{1,n}$

$$\begin{aligned} \hat{\psi}_n\left(X_n - \check{u}_{1,n} + \frac{t - X_n}{M_n - m}\right) &\geq \frac{1}{M_n - m} \hat{\psi}_n(t - \check{u}_{1,n}) + \frac{M_n - m - 1}{M_n - m} \hat{\psi}_n(X_n - \check{u}_{1,n}) \\ &\geq \frac{m}{M_n - m} + \frac{M_n(M_n - m - 1)}{M_n - m} = M_n - 1. \end{aligned}$$

Thus,

$$\text{Leb}(x \in \mathbb{R} : \hat{\psi}_n(x) \geq M_n - 1) \geq \text{Leb} \left(\frac{M_n - m - 1}{M_n - m} (X_n - \check{u}_{1,n}) + \frac{1}{M_n - m} \mathcal{C}_k \right) = \frac{|\mathcal{C}_k|}{M_n - m},$$

and hence

$$1 = \int_{\mathbb{R}} e^{\hat{\psi}_n(t)} dt \geq \frac{e^{M_n-1} |\mathcal{C}_k|}{M_n - m}$$

Hence, the event on the right side of (22) is included in

$$\left\{ m \leq M_n - e^{M_n-1} |\mathcal{C}_k| < M_n - e^{M_n-1} \eta \right\}.$$

Let $B_1 > 0$ be large enough so that $B_1 > B_0$, $B_1 - e^{B_1-1} \eta < -e^{B_1-1} \eta/2$ and $x \mapsto x - e^{x-1} \eta/2$ is decreasing on $[B_1, \infty)$, and consider the event

$$\left\{ M_n > B_1, |\mathcal{C}_k| > \eta \right\}.$$

Then, this event is included in

$$\left\{ m \leq -e^{M_n-1} \eta/2 \right\}$$

and occurrence of the latter implies in turn that

$$\Phi_n(\log(\hat{g}_n)) \leq \frac{1}{n} \left[([\alpha \bar{N}] - 1) M_n - (\bar{N} - [\alpha \bar{N}] + 1) \frac{1}{2} e^{M_n-1} \eta + \frac{n - \bar{N}}{n} M_n \right],$$

and consequently

$$\Phi_n(\log(\hat{g}_n)) \leq \left(1 - (1 - \alpha) \frac{\bar{N}}{n} \right) M_n - \frac{(1 - \alpha) \bar{N}}{2n} e^{M_n-1} \eta.$$

Using similar arguments as in the proof of our Proposition A.1, we can increase n_0 to ensure that for $n \geq n_0$

$$P \left(\frac{\bar{N}}{n} \geq \frac{G^0(\bar{u}^0)}{2} \right) \geq 1 - \epsilon/5.$$

Hence, occurrence of the event

$$\left\{ M_n > B_1, |\mathcal{C}_k| > \eta, \frac{\bar{N}}{n} \geq \frac{G^0(\bar{u}^0)}{2} \right\}.$$

implies that

$$\Phi_n(\log(\hat{g}_n)) \leq \left(1 - (1 - \alpha) \frac{G^0(\bar{u}^0)}{2} \right) M_n - \frac{(1 - \alpha) G^0(\bar{u}^0)}{4} \eta e^{M_n-1} \quad (23)$$

Next, if we take $e^w = \varphi_\sigma$ the density of standard Normal random variable. Then,

$$\begin{aligned}\Phi_n(\log(\hat{g}_n)) &\geq \Phi_n\left(\check{\pi}_n e^{w(\cdot - \check{u}_{1,n})} + (1 - \check{\pi}_n) e^{w(\cdot - \check{u}_{2,n})}\right) \\ &= \int_{\mathbb{R}} \log\left(\check{\pi}_n e^{w(x - \check{u}_{1,n})} + (1 - \check{\pi}_n) e^{w(x - \check{u}_{2,n})}\right) d(\mathbb{G}_n(x) - G^0(x)) \\ &\quad + \int_{\mathbb{R}} \log\left(\check{\pi}_n e^{w(x - \check{u}_{1,n})} + (1 - \check{\pi}_n) e^{w(x - \check{u}_{2,n})}\right) g^0(x) dx - 1.\end{aligned}$$

For $\eta > 0$ small enough, it follows from consistency of $\check{u}_{1,n}, \check{u}_{2,n}$ and $\check{\pi}_n$ that the event

$$\mathcal{E}^0 = \left\{ |\check{\pi}_n - \check{\pi}_n^0| \leq \eta/2, |\check{u}_{1,n} - u_1^0| \leq \eta/2, |\check{u}_{2,n} - u_2^0| \leq \eta/2 \right\}$$

occurs with probability greater than $1 - \epsilon/5$ for $n \geq n_0$ (at the cost of increasing the previous n_0). By Proposition A.3, since (f^0, e^w) certainly satisfy Condition (C), this implies that the event

$$\mathcal{E}_{\text{emp}} = \left\{ \left| \int_{\mathbb{R}} \log\left(\check{\pi}_n e^{w(x - \check{u}_{1,n})} + (1 - \check{\pi}_n) e^{w(x - \check{u}_{2,n})}\right) d(\mathbb{G}_n(x) - G^0(x)) \right| \leq \frac{1}{2} \right\}$$

occurs with probability greater than $1 - \epsilon/5$ for $n \geq n_0$. Also, there exists a constant $D > 0$ (depending on ϵ) such that the event

$$\mathcal{E}_{KL} = \left\{ \left| \int_{\mathbb{R}} \log\left(\check{\pi}_n e^{w(x - \check{u}_{1,n})} + (1 - \check{\pi}_n) e^{w(x - \check{u}_{2,n})}\right) g^0(x) dx \right| \leq D \right\}$$

with probability greater than $1 - \epsilon/5$. Indeed, note that $\check{\pi}_n e^{w(x - \check{u}_{1,n})} + (1 - \check{\pi}_n) e^{w(x - \check{u}_{2,n})} \leq 1$ for all $x \in \mathbb{R}$ so that

$$\begin{aligned}&\left| \int_{\mathbb{R}} \log\left(\check{\pi}_n e^{w(x - \check{u}_{1,n})} + (1 - \check{\pi}_n) e^{w(x - \check{u}_{2,n})}\right) g^0(x) dx \right| \\ &= - \int_{\mathbb{R}} \log\left(\check{\pi}_n e^{w(x - \check{u}_{1,n})} + (1 - \check{\pi}_n) e^{w(x - \check{u}_{2,n})}\right) g^0(x) dx.\end{aligned}$$

Using this fact and concavity of the logarithm, it follows that

$$\begin{aligned}0 &\leq - \int_{\mathbb{R}} \log\left(\check{\pi}_n e^{w(x - \check{u}_{1,n})} + (1 - \check{\pi}_n) e^{w(x - \check{u}_{2,n})}\right) g^0(x) dx \\ &\leq \frac{1}{2} \int_{\mathbb{R}} (\check{\pi}_n (x - \check{u}_{1,n})^2 + (1 - \check{\pi}_n) (x - \check{u}_{2,n})^2) g^0(x) dx \\ &< \infty\end{aligned}$$

since log-concave densities have finite moments of any order. Put $L = -3/2 - D$, and let $B_2 > B_1$ such that

$$\left(1 - (1 - \alpha) \frac{G(\bar{u}^0)}{2}\right) B_2 - \frac{(1 - \alpha) G(\bar{u}^0)}{4} \eta e^{B_2 - 1} < L$$

and the function $x \mapsto \left(1 - (1 - \alpha) \frac{G(\bar{u}^0)}{2}\right) x - \frac{(1 - \alpha)G(\bar{u}^0)}{4} \eta e^{x-1}$ is decreasing on $[B_2, \infty)$. By the choice of B_2 above, we see that

$$\mathcal{E}_\cap = \left\{ M_n > B_2, |C_k| > \eta, \frac{\bar{N}}{n} \geq \frac{G^0(\bar{u}^0)}{2} \right\} \cap \mathcal{E}^0 \cap \mathcal{E}_{\text{emp}} \cap \mathcal{E}_{KL} = \emptyset$$

Hence,

$$P\left(\|\hat{g}_n\|_\infty > B_2\right) = P(M_n > B_2) = P(M_n > B_2, \mathcal{E}_\cap^c) \leq P(\mathcal{E}_\cap^c) \leq \epsilon.$$

Note that B_2 depends on ϵ through L . □

To establish the first consistency result, note by the definition of \hat{g}_n we have that for any density \tilde{g} of the form

$$\tilde{g} = \check{\pi}_n f(\cdot - \check{u}_{1,n}) + (1 - \check{\pi}_n) f(\cdot - \check{u}_{2,n}).$$

for $f \in \mathcal{SLC}_1$ that

$$0 \leq \int \log \hat{g}_n d\mathbb{G}_n - \int \log \tilde{g} d\mathbb{G}_n \leq \int \log(\hat{g}_n + b) d\mathbb{G}_n - \int \log \tilde{g} d\mathbb{G}_n.$$

Now let $b > 0$, and G^0 be the cdf of the true mixed density g^0 . As first established by [Pal et al. \(2007\)](#) in their Lemma 1, the inequality above yields

$$\begin{aligned} 2H^2(\hat{g}_n, g^0) &\leq \epsilon(b) + \int_{\mathbb{R}} \log(\hat{g}_n(t) + b) d(\mathbb{G}_n(t) - G^0(t)) \\ &\quad + \int_{\mathbb{R}} \log(g^0(t) + b) dG^0(t) - \int_{\mathbb{R}} \log(\tilde{g}(t)) dG^0(t) \\ &\quad - \int_{\mathbb{R}} \log(\tilde{g}(t)) d(\mathbb{G}_n(t) - G^0(t)) \end{aligned} \tag{24}$$

where $H(p, q)$ denotes the Hellinger distance and

$$\epsilon(b) = 2 \int_{\mathbb{R}} \sqrt{\frac{b}{b + g^0(t)}} dG^0(t).$$

The main idea behind introducing the small positive quantity b is to avoid integration issues due to the fact that $\log(\hat{g}_n) = -\infty$ outside a certain interval. In the problem of estimating a log-concave density, the empirical term on the right side of the previous inequality was shown to converge to zero using the fact that the maximum value of the log-concave MLE stays bounded in n , see Theorem 3.2 of [Pal et al. \(2007\)](#) and Lemma 4 in [Schuhmacher and Dümbgen \(2010\)](#) for a generalization of the same result in the multivariate setting. This property of the MLE was then combined with the fact that a level set of a bounded unimodal function is convex

and compact. This cannot be claimed anymore if we replace a unimodal function by a mixture of two unimodal functions, not even when those functions are log-concave. For this reason, we shall use instead the Glivenko-Cantelli results proved above.

In our proof, the natural choice for \tilde{g} would be

$$\tilde{g}_n = \tilde{\pi}_n f^0(\cdot - \tilde{u}_{1,n}) + (1 - \tilde{\pi}_n) f^0(\cdot - \tilde{u}_{2,n}), \quad (25)$$

but this is problematic. For instance $\int \log \tilde{g}_n d\mathbb{G}_n$ can be $-\infty$ for arbitrarily large n if g^0 has compact support. This might occur when the smallest/ largest order statistic is very close to the left/right end of the support if f^0 and $\tilde{u}_{1,n}/\tilde{u}_{2,n}$ is smaller/larger than $\tilde{u}_{1,n}^0/u_2^0$. But even if g^0 does not have compact support, if it does have too large a slope (e.g., it approximates having compact support by dropping towards 0 very quickly at some point), then $\int \log \tilde{g}_n dG^0$ can still be infinite. Thus, we consider the surrogate density function

$$\tilde{g}_{n,h} = \tilde{\pi}_n f_h^0(\cdot - \tilde{u}_{1,n}) + (1 - \tilde{\pi}_n) f_h^0(\cdot - \tilde{u}_{2,n}) \quad (26)$$

where $f_h^0 = f^0 \star \phi_h$, the convolution of f^0 and the density of centered normal with standard deviation $h > 0$. This choice alleviates the above-mentioned problems; the risk of having a divergent log-likelihood is now excluded since f_h^0 , the component density of $\tilde{g}_{n,h}$, is supported on \mathbb{R} . Note that f_h^0 is a log-concave density by preservation of log-concavity under convolution ([Ibragimov, 1956](#)), and is symmetric.

Proof of Theorem 3.1: The proof starts from the inequality in (24) with $\tilde{g}_{n,h}$ (defined in (26)) as \tilde{g} . Theorem A.2 says that $\|\hat{g}_n\|_\infty$ is $O_p(1)$, so $\|\hat{f}_n\|_\infty$ is also $O_p(1)$, meaning that for any $\epsilon > 0$, there is an M such that \hat{f}_n lies in \mathcal{LC}_M (the class of log-concave functions bounded by M) with probability $1 - \epsilon$. Thus, we can apply Proposition A.2 to conclude that

$$\left| \int_{\mathbb{R}} \log(\hat{g}_n(t) + b) d(\mathbb{G}_n(t) - G^0(t)) \right| \leq \sup_{p \in \mathcal{F}_b} \left| \int_{\mathbb{R}} p d(\mathbb{G}_n - G^0) \right| < \epsilon$$

with probability $1 - \epsilon$ for n large. Now, let us write

$$g_h^0(t) = \tilde{\pi}_n^0 f^0(\cdot - u_1^0) \star \phi_h + (1 - \tilde{\pi}_n^0) f^0(\cdot - u_2^0) \star \phi_h$$

where ϕ_h is the density of a $N(0, h^2)$ random variable. It then follows from Lemma A.2 that as $n \rightarrow \infty$,

$$\int_{\mathbb{R}} \log(\tilde{g}_{n,h}) dG^0 \rightarrow_p \int_{\mathbb{R}} \log(g_h^0) dG^0.$$

Now, since $\int_{\mathbb{R}} |\log(g^0)| dG^0 < \infty$, by Lemma 3.4 of [Seregin and Wellner \(2010\)](#),

$$\epsilon(b) + \int (\log(g^0(t) + b) - \log(g_h^0(t))) dG^0(t) \rightarrow \int (\log(g^0) - \log(g_h^0)) dG^0 \quad (27)$$

as $b \searrow 0$, by the dominated convergence theorem. Now, for $|t|$ large enough, $g_h^0(t) \geq g^0(t)$, so $|\log g_h^0(t)| \leq |\log g^0(t)|$, and by [Seregin and Wellner \(2010\)](#), as was just

mentioned, the latter is g^0 -integrable, so that the right side of (27) converges to 0 as $h \searrow 0$. Thus, for $\epsilon > 0$, we can take b and h small enough that

$$\left| \epsilon(b) + \int_{\mathbb{R}} \log(g^0(t) + b) dG^0(t) - \int_{\mathbb{R}} \log(\check{g}_{n,h}) dG^0 \right| < \epsilon$$

with high probability for n large enough. Now, with h fixed, by Proposition A.3,

$$\left| \int_{\mathbb{R}} \log(\check{g}_{n,h}) d(\mathbb{G}_n - G^0) \right| \leq \sup_{p \in \mathcal{H}} \left| \int_{\mathbb{R}} p d(\mathbb{G}_n - G^0) \right| < \epsilon$$

with high probability for n large, where \mathcal{H} is defined as in Proposition A.3 with q taken as f_h^0 and p_0 taken as g^0 ; the proof of Lemma A.2 shows that since (f^0, ϕ_h) satisfy Condition (C) then $\log f_h^0$ is g_0 -integrable, and in fact for any $d \in \mathbb{R}$, $\log f_h^0(\cdot - d)$ is g_0 -integrable (and so is f_0 -integrable, so (f^0, f_h^0) satisfy the condition (C) as needed for Proposition A.3). Thus by (24) we are done. \square

Proof of Corollary 3.1. Recall the definition of \check{g}_n from (25), as well as the well-known fact that for any two densities p_1 and p_2 ,

$$\frac{1}{4} \left(\int_{\mathbb{R}} |p_1(t) - p_2(t)| dt \right)^2 \leq H^2(p_1, p_2) \leq \int_{\mathbb{R}} |p_1(t) - p_2(t)| dt. \quad (28)$$

Assuming that $\check{\pi}_n^0 < 1/2$ implies by consistency that $\check{\pi}_n < 1/2$ with increasing probability. Hence, the inversion formula (9) in Bordes et al. (2006) yields

$$\begin{aligned} \int_{\mathbb{R}} |\hat{f}_n(t) - f^0(t)| dt &\leq \frac{1}{1 - \check{\pi}_n} \sum_{k=0}^{\infty} \left(\frac{\check{\pi}_n}{1 - \check{\pi}_n} \right)^k \int_{\mathbb{R}} |\hat{g}_n(t) - \check{g}_n(t)| dt \\ &= \frac{1}{1 - 2\check{\pi}_n} \int_{\mathbb{R}} |\hat{g}_n(t) - \check{g}_n(t)| dt. \end{aligned} \quad (29)$$

From the first inequality in (28) and Theorem 3.1 above, it follows that

$$\int_{\mathbb{R}} |\hat{g}_n(t) - g^0(t)| dt = o_p(1)$$

which in turn implies that

$$\int_{\mathbb{R}} |\hat{g}_n(t) - \check{g}_n(t)| dt = o_p(1).$$

Indeed,

$$\int_{\mathbb{R}} |\hat{g}_n(t) - \check{g}_n(t)| dt \leq \int_{\mathbb{R}} |\hat{g}_n(t) - g^0(t)| dt + \int_{\mathbb{R}} |g^0(t) - \check{g}_n(t)| dt$$

with

$$\begin{aligned} \int_{\mathbb{R}} |g^0(t) - \check{g}_n(t)| dt &\leq 2|\check{\pi}_n - \check{\pi}_n^0| + \int_{\mathbb{R}} |f^0(t - \check{u}_{1,n}) - f^0(t - u_1^0)| dt \\ &\quad + \int_{\mathbb{R}} |f^0(t - \check{u}_{2,n}) - f^0(t - u_2^0)| dt \\ &= o_p(1) \end{aligned}$$

by consistency of $\tilde{\pi}_n$, $\tilde{u}_{1,n}$ and $\tilde{u}_{2,n}$ and Lemma A.3. It follows now from (29) and the inequalities in (28) that

$$H(\hat{f}_n, f^0) = o_p(1), \text{ and } \int_{\mathbb{R}} |\hat{f}_n(t) - f^0(t)| dt = o_p(1).$$

Note that convergence of the MLE \hat{f}_n in probability to 0 in the L_1 distance implies its weak convergence to f^0 with increasing probability. Hence, for any arbitrary sequence (n') we can extract a further subsequence (n'') such that $\hat{f}_{n''}$ converges weakly to f^0 almost surely. Hence, the assertion (c) in Proposition 2 of Cule and Samworth (2010) holds almost surely for $f_n = \hat{f}_{n''}$ and $f = f^0$, and the remaining claims of our lemma now follow since (n') was chosen arbitrarily. \square

Next, we prove the results in Section 4.

Proof of Proposition 4.1: It follows from the recent result of Doss and Wellner (2015); see their Theorem 4.1 for $s = 0$, that the classes

$$\left\{ f(\cdot - a), f \in \mathcal{S}\mathcal{L}\mathcal{C}, f(0) \in [1/M, M], a \in [a_0 - \delta, a_0 + \delta] \right\}$$

and

$$\left\{ f(\cdot - b), f \in \mathcal{S}\mathcal{L}\mathcal{C}, f(0) \in [1/M, M], b \in [b_0 - \delta, b_0 + \delta] \right\}$$

have both the same bracketing entropy $\log(N) \lesssim 1/\sqrt{\epsilon}$, where \lesssim means that the term on the left side is smaller or equal than the term on the right side up to a positive constant. Let $\{\epsilon, \dots, K\epsilon\}$, a ϵ -net for $[0, 1]$, where $K = \lfloor 1/\epsilon \rfloor + 1$. For $i \in \{1, \dots, N\}$, let $[l_i, u_i]$ and $[l'_i, u'_i]$ an ϵ -bracket for the first and second class respectively. Note that since the brackets are in the Hellinger sense, we have that $u_i \geq l_i \geq 0$ and $u'_i \geq l'_i \geq 0$. Now, there exist $i \in \{1, \dots, N\}$ and $j \in \{1, \dots, K\}$ such that

$$\underline{g} \equiv (j-1)\epsilon l_i + (1-j\epsilon)l'_i \leq \lambda f(x-a) + (1-\lambda)f(x-b) \leq \bar{g} \equiv j\epsilon u_i + (1-(j-1)\epsilon)u'_i$$

with

$$\begin{aligned} & H^2(\underline{g}, \bar{g}) \\ &= \frac{1}{2} \int_{\mathbb{R}} \left\{ \left(j\epsilon u_i(t) + (1-(j-1)\epsilon)u'_i(t) \right)^{1/2} - \left((j-1)\epsilon l_i(t) + (1-j\epsilon)l'_i(t) \right)^{1/2} \right\}^2 dt \\ &\leq 2H^2\left((j-1)\epsilon l_i + (1-(j-1)\epsilon)u'_i, j\epsilon u_i + (1-(j-1)\epsilon)u'_i \right) \\ &\quad + 2H^2\left((1-j\epsilon)l'_i + (j-1)\epsilon l_i, (1-(j-1)\epsilon)u'_i + (j-1)\epsilon l_i \right) \\ &\leq 2H^2\left((j-1)\epsilon l_i, j\epsilon u_i \right) + 2H^2\left((1-j\epsilon)l'_i, (1-(j-1)\epsilon)u'_i \right), \text{ by Lemma A.4} \\ &\leq 4(j-1)\epsilon H^2(l_i, u_i) + 4\left(\sqrt{j} - \sqrt{j-1} \right)^2 \epsilon \\ &\quad + 4(1-j\epsilon) H^2(l'_i, u'_i) + 4\left(\sqrt{1-(j-1)\epsilon} - \sqrt{1-j\epsilon} \right)^2, \text{ applying again Lemma A.4.} \end{aligned}$$

Using the fact that $0 \leq j\epsilon \leq 1$ and $1 - (j-1)\epsilon \geq \epsilon$ for all $j \in \{1, \dots, K\}$, we conclude from the preceding calculations that

$$H^2(\underline{g}, \bar{g}) \leq 4H^2(l_i, u_i) + 4H^2(l'_i, u'_i) + 8\epsilon \lesssim \epsilon.$$

The proof is complete by noting that

$$\begin{aligned} \log N_{[\cdot]}(\epsilon, \mathcal{G}, H) &\leq \log K + \log N \leq \log \left(\frac{1}{\epsilon} + 1 \right) + \log N \\ &\leq \frac{1}{\sqrt{\epsilon}} + \log N \lesssim \frac{1}{\sqrt{\epsilon}} \end{aligned}$$

using the fact that $\log(x+1) \leq \sqrt{x}$ for all $x \in [0, \infty)$. \square

To prepare for the proof of Theorem 4.1, we recall that $\tilde{\pi}_n$, $\tilde{u}_{1,n}$ and $\tilde{u}_{2,n}$ are estimates of $\tilde{\pi}_n^0$, u_1^0 and u_2^0 respectively, that are converging at the rate $1/\sqrt{n}$. Recall also that $\tilde{g}_n = \tilde{\pi}_n f^0(\cdot - \tilde{u}_{1,n}) + (1 - \tilde{\pi}_n) f^0(\cdot - \tilde{u}_{2,n})$, and that $\hat{g}_n = \tilde{\pi}_n \hat{f}_n(\cdot - \tilde{u}_{1,n}) + (1 - \tilde{\pi}_n) \hat{f}_n(\cdot - \tilde{u}_{2,n})$, where \hat{f}_n is the log-concave MLE. As done in [van der Vaart and Wellner \(1996\)](#) (page 326, Section 3.4), we consider the criterion function

$$r_{n,g} = \log \frac{g + \tilde{g}_n}{2\tilde{g}_n}$$

for g and

$$\tilde{g}_n = \tilde{\pi}_n \tilde{f}_n(\cdot - \tilde{u}_{1,n}) + (1 - \tilde{\pi}_n) \tilde{f}_n(\cdot - \tilde{u}_{2,n})$$

with $\tilde{f}_n \in \mathcal{SCG}_1$ to be constructed. Note that

$$r_{n,\tilde{g}_n} = 0, \text{ and } \mathbb{P}_n r_{n,\hat{g}_n} \geq \frac{1}{2} \mathbb{P}_n \log \frac{\hat{g}_n}{\tilde{g}_n} \geq 0 \quad (30)$$

where the second claim follows from the definition of the MLE \hat{g}_n and concavity of the logarithm.

Consider now the class of functions

$$\mathcal{R}_{n,\eta} = \left\{ r_{n,g} - r_{n,\tilde{g}_n} : g \in \mathcal{G}, H(g, \tilde{g}_n) < \eta \right\} = \left\{ r_{n,g} : g \in \mathcal{G}, H(g, \tilde{g}_n) < \eta \right\}.$$

If P^0 denotes again the true probability measure associated with g^0 , let

$$\mathbb{G}_n(r_{n,g}) = \sqrt{n} (\mathbb{P}_n - P^0) r_{n,g} = \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n r_{n,g}(X_i) - \int r_{n,g}(x) dP^0(x) \right)$$

and denote

$$\|\mathbb{G}_n\|_{\mathcal{R}_{n,\eta}} = \sup_{g \in \mathcal{R}_{n,\eta}} |\mathbb{G}_n(r_{n,g})|.$$

Finally, define

$$\tilde{J}_{[\cdot]}(\delta, \mathcal{G}, H) = \int_0^\delta \sqrt{1 + \log N_{[\cdot]}(\epsilon, \mathcal{G}, H)} d\epsilon.$$

Theorem 3.4.4 of [van der Vaart and Wellner \(1996\)](#) gives sufficient conditions to obtain control on $\|\mathbb{G}_n\|_{\mathcal{R}_{n,\eta}}$ in the mean. This control will involve the bracketing entropy bound obtained for the class \mathcal{G} . One of the crucial conditions to be fulfilled is that the sequence of densities \tilde{g}_n need to be chosen such that \tilde{g}_n approximates the truth g^0 and

$$\frac{g^0}{\tilde{g}_n} \leq M \tag{31}$$

for some $M > 0$. Note that the reason we cannot choose $\tilde{g}_n = g^0$ is that in this problem maximization of the log-likelihood involve the random variables $\check{\pi}_n, \check{u}_{1,n}$ and $\check{u}_{2,n}$, hence it is not at all straightforward to compare the values taken by the criterion at \hat{g}_n and g^0 . As will be shown in Proposition [A.4](#), we will exhibit an approximating sequence \tilde{g}_n that will satisfy the condition in (31) with increasing probability. Based on this proposition, we give now the proof of Theorem [4.1](#) along the same lines of the proof of Theorem 3.25 in [van der Vaart and Wellner \(1996\)](#); see page 290.

Proof of Theorem [4.1](#). Fix $\epsilon > 0$. Let \tilde{g}_n be the approximating sequence of Proposition [A.4](#), and consider the shells

$$S_{j,n} = \left\{ g : 2^{j-1} < n^{2/5} H(g, \tilde{g}_n) \leq 2^j \right\}$$

for integers $j \geq 1$. Fix an integer $J \geq 1$, and consider the event

$$\left\{ n^{2/5} H(\hat{g}_n, \tilde{g}_n) > 2^J \right\}.$$

Occurrence of this event implies that \hat{g}_n belongs to some $S_{j,n}$ with $j \geq J$. But our remark in (30) implies that

$$\sup_{g \in S_{j,n}} (\mathbb{P}_n r_{n,g} - \mathbb{P}_n r_{n,\tilde{g}_n}) = \sup_{g \in S_{j,n}} \mathbb{P}_n r_{n,g} \geq 0.$$

Thus, for any $\delta > 0$ and $M > 0$ we can write

$$\begin{aligned} P \left(n^{2/5} H(\hat{g}_n, \tilde{g}_n) > 2^J \right) &= P \left(n^{2/5} H(\hat{g}_n, \tilde{g}_n) > 2^J, g^0 \leq M \tilde{g}_n \right) + P \left(g^0 > M \tilde{g}_n \right) \\ &= \sum_{j \geq J, 2^j \leq n^{2/5} \eta} P \left(\sup_{g \in S_{j,n}} (\mathbb{P}_n r_{n,g} - \mathbb{P}_n r_{n,\tilde{g}_n}) \geq 0, g^0 \leq M \tilde{g}_n \right) \\ &\quad + P \left(H(\hat{g}_n, \tilde{g}_n) \geq \delta/2 \right) + P \left(g^0 > M \tilde{g}_n \right). \end{aligned}$$

By Proposition A.4, for any $\eta > 0$ and M large enough the second and third terms are bounded by $\epsilon/2$. Now, using Theorem 3.4.4 of [van der Vaart and Wellner \(1996\)](#), we have

$$P^0(r_{n,g} - r_{n,\tilde{g}_n}) \lesssim -H^2(g, \tilde{g}_n) \quad (32)$$

for all densities g such that $H(g, \tilde{g}_n) \geq 32MH(\tilde{g}_n, g^0)$, and

$$E_{P^0} \|\mathbb{G}_n\|_{\mathcal{R}_{n,\eta}} \lesssim \tilde{J}_{[\]}(\eta, \mathcal{G}, H) \left(1 + \frac{\tilde{J}_{[\]}(\eta, \mathcal{G}, H)}{\eta^2 \sqrt{n}} \right). \quad (33)$$

From Proposition 4.1, we know that there exists a constant $K > 0$ (not depending on n) such that

$$\log N_{[\]}(\epsilon, \mathcal{G}, H) \leq \frac{K}{\sqrt{\epsilon}}.$$

For $\eta > 0$ small enough, $1 < K/\sqrt{\epsilon}$ and hence

$$\tilde{J}_{[\]}(\eta, \mathcal{G}, H) \leq \int_0^\eta \frac{\sqrt{2K}}{\epsilon^{1/4}} d\epsilon = \frac{4\sqrt{2K}}{3} \eta^{3/4}.$$

Now, define

$$\phi_n(\eta) := \frac{4\sqrt{2K}}{3} \eta^{3/4} \left(1 + \frac{4\sqrt{2K}}{3\eta^{5/4}\sqrt{n}} \right).$$

We have that

$$\frac{\phi_n(\eta)}{\eta} \propto \frac{1}{\eta^{1/4}} \left(1 + \frac{4\sqrt{2K}}{3\eta^{5/4}\sqrt{n}} \right)$$

and hence the function $\eta \mapsto \phi_n(\eta)/\eta$ is decreasing on $(0, \infty)$. Also, if we put $r_n = n^{2/5}$ then,

$$\begin{aligned} r_n^2 \phi_n \left(\frac{1}{r_n} \right) &= n^{4/5} \frac{4\sqrt{2K}}{3} n^{-3/10} \left(1 + \frac{4\sqrt{2K}}{3} \right) \\ &= \frac{4\sqrt{2K}}{3} \left(1 + \frac{4\sqrt{2K}}{3} \right) \sqrt{n}. \end{aligned}$$

Now note that on the event $\{g^0 \leq M\tilde{g}_n\}$ it follows from (32) that for all $g \in S_{n,j}$

$$P^0(r_{n,g} - r_{n,\tilde{g}_n}) \leq \frac{-2^{2j-2}}{r_n^2}$$

and hence

$$\begin{aligned}
& \sum_{j \geq J, 2^j \leq n^{2/5} \eta} P \left(\sup_{g \in S_{j,n}} (\mathbb{P}_n r_{n,g} - \mathbb{P}_n r_{n,\tilde{g}_n}) \geq 0, g^0 \leq M \tilde{g}_n \right) \\
& \leq \sum_{j \geq J, 2^j \leq n^{2/5} \eta} P \left(\|\mathbb{G}_n\|_{S_{n,j}} \geq \frac{-2^{2j-2} \sqrt{n}}{r_n^2} \right) \\
& \leq \sum_{j \geq J} \frac{\phi_n(2^j/r_n) r_n^2}{2^{2j-2} \sqrt{n}}, \text{ by Markov's inequality} \\
& \leq \frac{1}{4} \sum_{j \geq J} \frac{\phi_n(1/r_n) r_n^2}{2^j \sqrt{n}}, \text{ using the fact that } \frac{\phi_n(ct)}{ct} \leq \frac{\phi_n(t)}{t} \text{ for } c > 1 \\
& \lesssim \sum_{j \geq J} 2^{-j} \rightarrow 0 \text{ as } J \rightarrow \infty
\end{aligned} \tag{34}$$

which implies that $H(\hat{g}_n, \tilde{g}_n) = O_p(n^{-2/5})$. Hence, $L_1(\hat{g}_n, \tilde{g}_n) = O_p(n^{-2/5})$ which in turn implies that $L_1(\hat{g}_n, g^0) = O_p(n^{-2/5})$ using the result of Proposition A.4 and the triangular inequality.

To conclude a similar rate result for \hat{f}_n , we show now that $L_1(\hat{f}_n, c_n \tilde{f}_n) = O_p(n^{-2/5})$ where \tilde{f}_n and the normalizing constant c_n are defined in Proposition A.4. Using again the inversion formula (9) in Bordes et al. (2006), we can write

$$\int_{\mathbb{R}} |\hat{f}_n(t) - c_n \tilde{f}_n(t)| dt \leq \frac{1}{1 - 2\tilde{\pi}_n} \int_{\mathbb{R}} |\hat{g}_n(t) - \tilde{g}_n(t)| dt = O_p(n^{-2/5}).$$

The proof is complete using the result of Proposition A.4 and the triangle inequality.

□

Proposition A.4 *There exist $\tilde{f}_n \in \mathcal{SCG}$ and $c_n > 0$ such that $c_n \tilde{f}_n \in \mathcal{SCG}_1$ and*

$$\sup_{t \in \mathbb{R}} \frac{g^0(t)}{\tilde{g}_n(t)} = O_p(1)$$

where

$$\tilde{g}_n = \tilde{\pi}_n c_n \tilde{f}_n(\cdot - \tilde{u}_{1,n}) + (1 - \tilde{\pi}_n) c_n \tilde{f}_n(\cdot - \tilde{u}_{2,n}).$$

Furthermore, if $\sqrt{n}(\tilde{u}_{1,n} - u_1^0) = O_p(1)$, $\sqrt{n}(\tilde{u}_{2,n} - u_2^0) = O_p(1)$ and $\sqrt{n}(\tilde{\pi}_n - \pi_n^0) = O_p(1/\sqrt{n})$, then

$$L_1(\tilde{f}_n, f^0) = O_p(n^{-1/2}), \text{ and } L_1(\tilde{g}_n, g^0) = O_p(n^{-1/2}).$$

Proof of Proposition A.4. In the following we denote

$$\delta_1 = \tilde{u}_{1,n} - u_1^0 \text{ and } \delta_2 = \tilde{u}_{2,n} - u_2^0.$$

Suppose that $0 \leq \delta_2 \leq \delta_1$. Define

$$\tilde{f}_n(t) = \begin{cases} f^0(|t| - \delta_1), & \text{if } t \leq -\delta_1 \\ f^0(0), & \text{if } -\delta_1 \leq t \leq \delta_1 \\ f^0(t - \delta_1), & \text{if } t \geq \delta_1. \end{cases}$$

It is not difficult to see that $\tilde{f}_n \in \mathcal{SLC}$. Now, define the ratios

$$\tilde{R}_{n,1}(t) := \frac{f^0(t)}{\tilde{f}_n(t - \delta_1)}, \quad \text{and} \quad \tilde{R}_{n,2}(t) := \frac{f^0(t)}{\tilde{f}_n(t - \delta_2)}.$$

We have that

$$\tilde{R}_{n,1}(t) = \begin{cases} \frac{f^0(t)}{f^0(\delta_1 - t - \delta_1)} = 1, & \text{if } t \leq 0 \\ \frac{f^0(t)}{f^0(0)} \leq 1, & \text{if } 0 \leq t \leq 2\delta_1 \\ \frac{f^0(t)}{f^0(t - 2\delta_1)} \leq 1, & \text{if } t \geq 2\delta_1. \end{cases}$$

The third inequality follows from the fact that f^0 is decreasing on $[0, \infty)$. Also,

$$\tilde{R}_{n,2}(t) = \begin{cases} \frac{f^0(t)}{f^0(\delta_2 - \delta_1 - t)} \leq 1, & \text{if } t \leq \delta_2 - \delta_1 \\ \frac{f^0(t)}{f^0(0)} \leq 1, & \text{if } \delta_2 - \delta_1 \leq t \leq \delta_1 + \delta_2 \\ \frac{f^0(t)}{f^0(t - \delta_1 - \delta_2)} \leq 1, & \text{if } t \geq \delta_1 + \delta_2. \end{cases}$$

The first inequality follows from symmetry of f^0 which allows to write that $f^0(t)/f^0(\delta_2 - \delta_1 - t) = f^0(x)/f^0(x - (\delta_1 - \delta_2))$ with $x = -t \geq \delta_1 - \delta_2 \geq 0$. Since f^0 is decreasing on $[0, \infty)$, it follows that $f^0(x) \leq f^0(x - (\delta_1 - \delta_2))$. The third inequality is again a consequence of the latter property of f^0 .

Now let $c_n = \left(\int_{\mathbb{R}} \tilde{f}_n(t) dt \right)^{-1}$. We have that

$$\begin{aligned} c_n^{-1} - 1 &= \int_{-\infty}^{-\delta_1} f^0(|t| - \delta_1) dt + \int_{\delta_1}^{\infty} f^0(t - \delta_1) dt + 2\delta_1 f^0(0) - 1 \\ &= 2\delta_1 f^0(0). \end{aligned}$$

This in turn implies that

$$0 \leq 1 - c_n = \frac{2\delta_1 f^0(0)}{1 + 2\delta_1 f^0(0)}.$$

Now, define $\tilde{g}_n := c_n \left(\tilde{\pi}_n \tilde{f}_n(\cdot - \tilde{u}_{1,n}) + (1 - \tilde{\pi}_n) \tilde{f}_n(\cdot - \tilde{u}_{2,n}) \right)$. Then,

$$\begin{aligned} \frac{g^0(t)}{\tilde{g}_n(t)} &= c_n^{-1} \frac{\tilde{\pi}_n^0 f^0(t - u_1^0) + (1 - \tilde{\pi}_n^0) f^0(t - u_2^0)}{\tilde{\pi}_n \tilde{f}_n(t - \tilde{u}_{1,n}) + (1 - \tilde{\pi}_n) \tilde{f}_n(t - \tilde{u}_{2,n})} \\ &\leq c_n^{-1} \frac{\tilde{\pi}_n^0}{\tilde{\pi}_n} \frac{f^0(t - u_1^0)}{\tilde{f}_n(t - u_1^0 - \delta_1)} + c_n^{-1} \frac{1 - \tilde{\pi}_n^0}{1 - \tilde{\pi}_n} \frac{f^0(t - u_2^0)}{\tilde{f}_n(t - u_2^0 - \delta_2)} \\ &= c_n^{-1} \frac{\tilde{\pi}_n^0}{\tilde{\pi}_n} \tilde{R}_{n,1}(t - u_1^0) + c_n^{-1} \frac{1 - \tilde{\pi}_n^0}{1 - \tilde{\pi}_n} \tilde{R}_{n,2}(t - u_2^0) \\ &\leq c_n^{-1} \left(\frac{\tilde{\pi}_n^0}{\tilde{\pi}_n} + \frac{1 - \tilde{\pi}_n^0}{1 - \tilde{\pi}_n} \right). \end{aligned}$$

By consistency of $\tilde{\pi}_n$ and $\tilde{u}_{2,n}$, we have $\tilde{\pi}_n^0 < \tilde{\pi}_n \leq (3/2)\tilde{\pi}_n^0$ and $c_n^{-1} \leq 3/2$ with increasing probability. Hence, we can bound the right hand side of the preceding display by $9/2$ with increasing probability. As the other cases can be handled similarly, the details are skipped but we give below the corresponding expression of \tilde{f}_n :

- If $0 \leq \delta_1 < \delta_2$, then we only need to switch the roles of δ_1 and δ_2 and hence take

$$\tilde{f}_n(t) = \begin{cases} f^0(|t| - \delta_2), & \text{if } t \leq -\delta_2 \\ f^0(0), & \text{if } -\delta_2 \leq t \leq \delta_2 \\ f^0(t - \delta_2), & \text{if } t \geq \delta_2. \end{cases}$$

- If $\delta_2 \leq 0 \leq \delta_1$, then we can take

$$\tilde{f}_n(t) = \begin{cases} f^0(|t| - (\delta_1 - \delta_2)), & \text{if } t \leq -(\delta_1 - \delta_2) \\ f^0(0), & \text{if } -(\delta_1 - \delta_2) \leq t \leq \delta_1 - \delta_2 \\ f^0(t - (\delta_1 - \delta_2)), & \text{if } t \geq \delta_1 - \delta_2. \end{cases}$$

- If $\delta_1 \leq 0 \leq \delta_2 \leq 0$, then we only need to switch δ_1 and δ_2 .
- If $\delta_1 \leq \delta_2 \leq 0$, then we can take

$$\tilde{f}_n(t) = \begin{cases} f^0(|t| + \delta_1), & \text{if } t \leq \delta_1 \\ f^0(0), & \text{if } \delta_1 \leq t \leq -\delta_1 \\ f^0(t + \delta_1), & \text{if } t \geq -\delta_1. \end{cases}$$

- If $\delta_2 \leq \delta_1 \leq 0$, we again switch the roles of δ_1 and δ_2 .

In all the cases above, one can verify that the ratios $\tilde{R}_{n,1}$ and $\tilde{R}_{n,2}$ as defined above stay below 1. We would like to stress the fact that the way \tilde{f}_n is constructed is not unique: one only need to exhibit examples which would give control of the ratio g^0/\tilde{g}_n . To show now the second assertion, we will again consider only the first case where $0 \leq \delta_2 \leq \delta_1$ since the remaining configurations can be handled similarly. We have that

$$\begin{aligned} L_1(\tilde{f}_n, f^0) &= \int_{\mathbb{R}} |\tilde{f}_n(t) - f^0(t)| dt \\ &= \int_{-\infty}^{-\delta_1} |f^0(-t - \delta_1) - f^0(t)| dt + \int_{-\delta_1}^{\delta_1} |f^0(0) - f(t)| dt \\ &\quad + \int_{\delta_1}^{\infty} |f^0(t - \delta_1) - f^0(t)| dt \\ &= 2 \int_{\delta_1}^{\infty} |f^0(t - \delta_1) - f^0(t)| dt + 2\delta_1 f^0(0) \\ &\leq 2(C+1)\delta_1 \end{aligned}$$

where C is the constant given in Lemma A.3. By the assumption on the rate of convergence of δ_1 , this implies that $L_1(\tilde{f}_n, f^0) = O_p(n^{-1/2})$. Also, we have

$$\begin{aligned}
& L_1(\tilde{g}_n, g^0) \\
&= \int_{\mathbb{R}} |\tilde{g}_n(t) - g^0(t)| dt \\
&= \int_{\mathbb{R}} \left| c_n \left(\tilde{\pi}_n \tilde{f}_n(t - \tilde{u}_{1,n}) + (1 - \tilde{\pi}_n) \tilde{f}_n(t - \tilde{u}_{2,n}) \right) - (\tilde{\pi}_n^0 f^0(t - \tilde{u}_{1,n}^0) + (1 - \tilde{\pi}_n^0) f^0(t - u_2^0)) \right| dt \\
&\leq 1 - c_n + \int_{\mathbb{R}} \left| \tilde{\pi}_n \tilde{f}_n(t - \tilde{u}_{1,n}) - \tilde{\pi}_n^0 f^0(t - \tilde{u}_{1,n}^0) \right| dt \\
&+ \int_{\mathbb{R}} \left| (1 - \tilde{\pi}_n) \tilde{f}_n(t - \tilde{u}_{2,n}) - (1 - \tilde{\pi}_n^0) f^0(t - u_2^0) \right| dt \\
&= \leq 1 - c_n + 2|\tilde{\pi}_n - \tilde{\pi}_n^0| + \int_{\mathbb{R}} \left| \tilde{f}_n(t - \tilde{u}_{1,n}) - f^0(t - u_1^0) \right| dt + \int_{\mathbb{R}} \left| \tilde{f}_n(t - \tilde{u}_{2,n}) - f^0(t - u_2^0) \right| dt \\
&= \leq 1 - c_n + 2|\tilde{\pi}_n - \tilde{\pi}_n^0| + \int_{\mathbb{R}} \left| \tilde{f}_n(t - \delta_1) - f^0(t) \right| dt + \int_{\mathbb{R}} \left| \tilde{f}_n(t - \delta_2) - f^0(t) \right| dt.
\end{aligned}$$

Using the definition of \tilde{f}_n for the first case, $0 \leq \delta_2 < \delta_1$, we can write

$$\begin{aligned}
\int_{\mathbb{R}} \left| \tilde{f}_n(t - \delta_1) - f^0(t) \right| dt &= \int_{-\infty}^0 |f^0(\delta_1 - t - \delta_1) - f^0(t)| + \int_0^{2\delta_1} |f^0(0) - f^0(t)| dt \\
&+ \int_{2\delta_1}^{\infty} |f^0(t - 2\delta_1) - f^0(t)| dt \\
&\leq 2\delta_1 f^0(0) + 6\delta_1 f^0(0), \text{ using Lemma A.1} \\
&= 8\delta_1 f^0(0) \\
&= O_p(n^{-1/2})
\end{aligned}$$

using again Lemma A.3, and the assumption on the rate of convergence of $\tilde{\pi}_n$ and $\tilde{u}_{1,n}$ and $\tilde{u}_{2,n}$. □

A.2 Auxiliary Results

Lemma A.3 *Let $f \in S\mathcal{LC}_1$. Then, there exists a constant $C > 0$ depending only on f such that*

$$\int_{\mathbb{R}} |f(t + \delta) - f(t)| dt \leq C\delta$$

for all $|\delta| \leq 1$.

Proof. For $\delta \geq 0$, we have $f(t + \delta) \leq f(t)$ on $t \in [0, \infty)$, $f(t - \delta) \geq f(t)$ on

$t \in [\delta/2, \infty)$ and $f(t - \delta) \leq f(t)$ on $[0, \delta/2]$. Using the symmetry of f , we can write

$$\begin{aligned}
\int_{\mathbb{R}} |f(t + \delta) - f(t)| dt &= \int_0^\infty |f(t + \delta) - f(t)| dt + \int_0^\infty |f(t - \delta) - f(t)| dt \\
&= \int_0^\infty (f(t) - f(t + \delta)) dt + \int_0^{\delta/2} (f(t) - f(t - \delta)) dt \\
&\quad + \int_{\delta/2}^\infty (f(t - \delta) - f(t)) dt \\
&\leq 1/2 - \int_\delta^\infty f(t) dt + \int_0^{\delta/2} f(t) dt + \int_{-\delta/2}^\infty f(t) dt - \int_{\delta/2}^\infty f(t) dt \\
&= \int_0^\delta f(t) dt + \int_0^{\delta/2} f(t) dt + 2 \int_0^{\delta/2} f(t) dt \\
&\leq 3\delta \sup_{t \in [0, 1]} f(t) = 3\delta f(0)
\end{aligned}$$

provided that $\delta \leq 1$. The same reasoning can be applied for negative values of δ .
 \square

Lemma A.4 *For any positive functions p, q, h we have that*

$$H(p + h, q + h) \leq H(p, q)$$

Proof. By definition, we have that

$$\begin{aligned}
2H^2(p + h, q + h) &= \int_{\mathbb{R}} \left(\sqrt{p(t) + h(t)} - \sqrt{q(t) + h(t)} \right)^2 dt \\
&= \int_{\mathbb{R}} \left(\frac{p(t) - q(t)}{\sqrt{p(t) + q(t)} + \sqrt{q(t) + h(t)}} \right)^2 dt \\
&= \int_{\mathbb{R}} \left(\sqrt{p(t)} - \sqrt{q(t)} \right)^2 \left(\frac{\sqrt{p(t)} + \sqrt{q(t)}}{\sqrt{p(t) + h(t)} + \sqrt{q(t) + h(t)}} \right)^2 dt \\
&\leq \int_{\mathbb{R}} \left(\sqrt{p(t)} - \sqrt{q(t)} \right)^2 dt = 2H^2(p, q)
\end{aligned}$$

where the last inequality follows since $h \geq 0$, hence the result. \square

The function J and its partial derivatives: As in [Dümbgen and Rufibach \(2009\)](#), we consider the two-dimensional function J defined by

$$J(r, s) = \int_0^1 \exp((1 - t)r + ts) dt$$

for $r, s \in \mathbb{R}$. Using the same notation of these authors, define

$$J_{a,b}(r, s) = \frac{\partial^{a+b} J(r, s)}{\partial^a s \partial^b s}.$$

Direct calculations yield

$$J_{a,b}(r, s) = \exp(r) J_{a,b}(0, s - r)$$

with

$$J_{0,0}(0, y) = J(0, y) = \begin{cases} 1, & \text{if } y = 0 \\ \frac{\exp(y)-1}{y}, & \text{otherwise,} \end{cases}$$

$$J_{0,1}(0, y) = \begin{cases} \frac{1}{2}, & \text{if } y = 0 \\ \frac{y \exp(y) - (\exp(y)-1)}{y^2}, & \text{otherwise,} \end{cases}$$

and

$$J_{0,2}(0, y) = \begin{cases} \frac{1}{3}, & \text{if } y = 0 \\ \frac{y^2 \exp(y) - 2y \exp(y) + 2(\exp(y)-1)}{y^3}, & \text{otherwise.} \end{cases}$$

References

- AZZALINI, A. and BOWMAN, A. W. (1990). A look at some data on the old faithful geyser. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **39** 357–365.
- BALABDAOUI, F. and BUTUCEA, C. (2014). On location mixtures with Pólya frequency components. *Submitted*.
- BALABDAOUI, F., JANKOWSKI, H., RUFIBACH, K. and PAVLIDES, M. (2013). Asymptotics of the discrete log-concave maximum likelihood estimator and related applications. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, **75** 769–790. URL <http://dx.doi.org/10.1111/rssb.12011>.
- BALABDAOUI, F., RUFIBACH, K. and WELLNER, J. A. (2009). Limit distribution theory for maximum likelihood estimation of a log-concave density. *Ann. Statist.*, **37** 1299–1331.
- BHATTACHARYA, R. N. and RANGA RAO, R. (1976). *Normal approximation and asymptotic expansions*. John Wiley & Sons, New York-London-Sydney. Wiley Series in Probability and Mathematical Statistics.
- BORDES, L., MOTTELET, S. and VANDEKERKHOVE, P. (2006). Semiparametric estimation of a two-component mixture model. *Ann. Statist.*, **34** 1204–1232.

- BUTUCEA, C. and VANDEKERKHOVE, P. (2014). Semiparametric mixtures of symmetric distributions. *The Scandinavian Journal of Statistics*, **41** 227–239.
- CHANG, G. T. and WALTHER, G. (2007). Clustering with mixtures of log-concave distributions. *Computational Statistics & Data Analysis*, **51** 6242–6251.
- CHEE, C.-S. and WANG, Y. (2013). Estimation of finite mixtures with symmetric components. *Stat. Comput.*, **23** 233–249. URL <http://dx.doi.org/10.1007/s11222-011-9305-5>.
- CHEN, Y. and SAMWORTH, R. J. (2013). Smoothed log-concave maximum likelihood estimation with applications. *Statist. Sinica*, **23** 1373–1398.
- CIULLO, M. (2009). Genetic park of cilento and vallo di dano project. <http://www.igb.cnr.it/cilentoisolates>. Accessed: 2014-04-20.
- COLONNA, V., NUTILE, T., ASTORE, M., GUARDIOLA, O., ANTONIOL, G., CIULLO, M. and PERSICO, M. G. (2007). Campora: A young genetic isolate in south italy. *Human heredity*, **64** 123–135.
- CULE, M. and SAMWORTH, R. (2010). Theoretical properties of the log-concave maximum likelihood estimator of a multidimensional density. *Electronic J. Stat.*, **4** 254–270.
- CULE, M., SAMWORTH, R. and STEWART, M. (2010). Maximum likelihood estimation of a multidimensional log-concave density. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **72** 545–607.
- DEMPSTER, A. P., LAIRD, N. M. and RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B*, **39** 1–38. With discussion.
- DOSS, C. and WELLNER, J. A. (2015). Global rates of convergence of the MLEs of log-concave and s -concave densities. *Annals of Statistics*,. To appear.
- DÜMBGEN, L., HÜSLER, A. and RUFIBACH, K. (2010). Active set and EM algorithms for log-concave densities based on complete and censored data. Tech. rep., University of Bern. Available at arXiv:0707.4643.
- DÜMBGEN, L. and RUFIBACH, K. (2009). Maximum likelihood estimation of a log-concave density and its distribution function. *Bernoulli*, **15** 40–68.
- DÜMBGEN, L. and RUFIBACH, K. (2011). logcondens: Computations related to univariate log-concave density estimation. *Journal of Statistical Software*, **39** 1–28.
- DÜMBGEN, L., SAMWORTH, R. and SCHUHMACHER, D. (2011). Approximation by log-concave distributions with applications to regression. *Ann. Statist.*, **39** 702–730.

- DÜMBGEN, L., SCHUHMACHER, D. and SAMWORTH, R. (2010). Approximation by log-concave distributions with applications to regression. *arXiv:1002.3448v3*. [arXiv:1002.3448v3](https://arxiv.org/abs/1002.3448v3).
- EILERS, P. H. C. and BORGDORFF, M. W. (2007). Non-parametric log-concave mixtures. *Computational Statistics & Data Analysis*, **51** 5444–5451.
- EVERITT, B. S. and HAND, D. J. (1981). *Finite Mixture Distributions*. Chapman & Hall, London-New York.
- FRALEY, C. and RAFTERY, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association*, **97** 611–631.
- HUNTER, D. R., WANG, S. and HETTMANSPERGER, T. P. (2007). Inference for mixtures of symmetric distributions. *Ann. Statist.*, **35** 224–251. URL <http://dx.doi.org/10.1214/009053606000001118>.
- IBRAGIMOV, I. A. (1956). On the composition of unimodal distributions. *Teor. Veroyatnost. i Primenen.*, **1** 283–288.
- KHAS’MINSKII, R. Z. (1978). A lower bound on the risks of nonparametric estimates of densities in the uniform metric. *Theory of Probability and its Applications*, **23** 794–798.
- KIM, A. K. H. and SAMWORTH, R. J. (2014). Global rates of convergence in log-concave density estimation. Available at [arXiv.org:1404.2298v1](https://arxiv.org/abs/1404.2298v1).
- KOENKER, R. and MIZERA, I. (2010). Quasi-concave density estimation. *Ann. Statist.*, **38** 2998–3027.
- MCLACHLAN, G. and PEEL, D. (2000). *Finite mixture models*. Wiley Series in Probability and Statistics: Applied Probability and Statistics, Wiley-Interscience, New York.
- NICULESCU, C. and PERSSON, L.-E. (2006). *Convex Functions and their Applications: a Contemporary Approach*. Springer Science & Business Media.
- PAL, J. K., WOODROOFE, M. B. and MEYER, M. C. (2007). Estimating a Polya frequency function. In *Complex Datasets and Inverse Problems: Tomography, Networks, and Beyond*, vol. 54 of *IMS Lecture Notes-Monograph Series*. IMS, 239–249.
- ROCKAFELLAR, R. T. (1970). *Convex analysis*. Princeton Mathematical Series, No. 28, Princeton University Press.
- RUFIBACH, K. (2007). Computing maximum likelihood estimators of a log-concave density function. *J. Statist. Comp. Sim.*, **77** 561–574.

- SCHOENBERG, I. J. (1951). On Pólya frequency functions. I. The totally positive functions and their Laplace transforms. *J. Analyse Math.*, **1** 331–374.
- SCHUHMACHER, D. and DÜMBGEN, L. (2010). Consistency of multivariate log-concave density estimators. *Statist. Probab. Lett.*, **80** 376–380. URL <http://dx.doi.org/10.1016/j.spl.2009.11.013>.
- SEREGIN, A. and WELLNER, J. A. (2010). Nonparametric Estimation of Multivariate Convex-Transformed Densities. *Ann. Statist.*, **38** 3751–3781.
- SILVERMAN, B. W. (1982). On the estimation of a probability density function by the maximum penalized likelihood method. *Ann. Statist.*, **10** 795–810.
- STONE, C. (1980). Optimal rates of convergence for nonparametric estimators. *The Annals of Statistics* 1–14.
- TITTERINGTON, D. M., SMITH, A. F. M. and MAKOV, U. E. (1985). *Statistical Analysis of Finite Mixture Distributions*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, John Wiley & Sons, Ltd., Chichester.
- VAN DE GEER, S. A. (2000). *Empirical Processes in M-Estimation*. Cambridge Univ Pr.
- VAN DER VAART, A. and WELLNER, J. A. (2000). Preservation theorems for Glivenko-Cantelli and uniform Glivenko-Cantelli classes. In *High dimensional probability, II*. Birkhäuser Boston, 115–133.
- VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer Series in Statistics, Springer-Verlag, New York.
- WALTHER, G. (2002). Detecting the presence of mixing with multiscale maximum likelihood. *J. Amer. Statist. Assoc.*, **97** 508–513.
- WEISBERG, S. (2005). *Applied Linear Regression*. 3rd ed. Wiley Series in Probability and Statistics, Wiley-Interscience, Hoboken, NJ.